"*Anti*-Bayesian" Statistical Pattern Recognition

B. John Oommen[1] and Others[2]

[1]*Chancellor's Professor*, *Fellow: IEEE*; *Fellow: IAPR*
School of Computer Science, Carleton University, Canada.

[2]Thomas (*Blackberry, Carleton University*);
Yazidi and Hammer (*Oslo University College*);
Khoury and Schmidt (*Lakehead University*)

What is the Goal of this Research?

- Proposes an “Anti-Bayesian” Paradigm for Pattern Classification
- Presents formal analysis for many distributions within the Exponential family
- Demonstrates the strength of the method with experimental results
- Provides the rationale for Border Identification (BI) Algorithms
- Proposes how to use an “Anti-Bayesian” Paradigm for Clustering
- Proposes how to use an “Anti-Bayesian” Paradigm for Text Classification

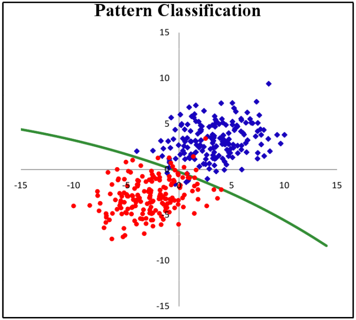
Relevant Background

- Pattern Classification
- Prototype Reduction Schemes
- Border Identification Algorithms

Pattern Classification

Categorize feature vectors into classes based on their features

- Bayesian classifier (Optimal)
- NN classifier
- Parzen window
- SVM
- Decision trees
- Etc...



Bayes’ Classifier

Bayes’ Thm: *A-posteriori* probability from *a-priori* probability

$$P(\omega_i/x) = \frac{p(x/\omega_i) \cdot P(\omega_i)}{p(x)}$$

Since denominator is common for all classes,

$\Longrightarrow$ Decide ω_i as class maximizing $P(\omega_i/x)$

$\Longrightarrow$ Minimizes classification error!!!

Problem with Bayes’ classifier

1. Training can take time
2. Must evaluate $p(\mathbf{x}^*/\omega_i)$ for testing sample
3. Can be very cumbersome
4. Attempt to simplify expression
5. Parametric *form* may be unknown

Solution: Bayes' classifier(Gaussian)

If $\sum_1 = \sum_2$,

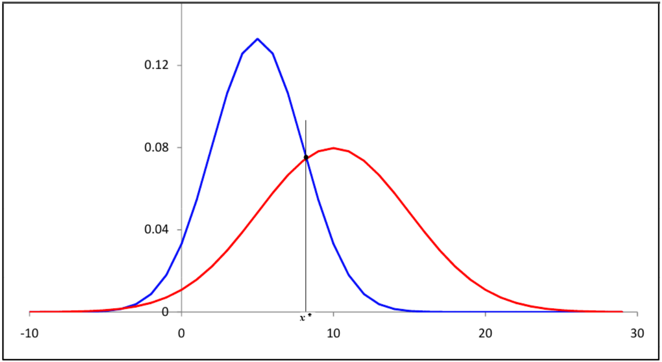
$$P(\omega_1|X) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} P(\omega_2|X)$$

$$\Longrightarrow p(X|\omega_1)P(\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(X|\omega_2)P(\omega_2)$$

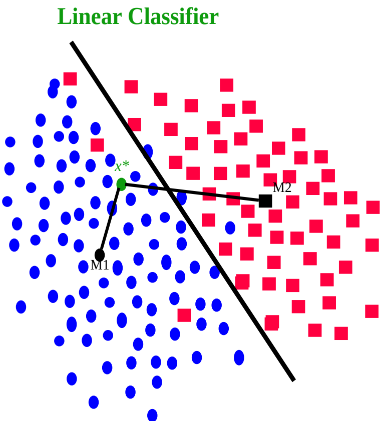
$$\Longrightarrow \|X - M_1\| \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \|X - M_2\|$$

Solution: Compare sample to Class Means!

Bayes’ Classifier for Uni-dimensional Distributions

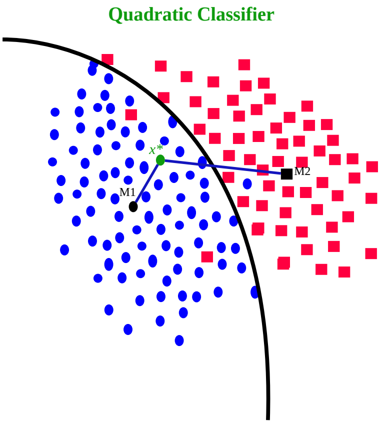


Linear Classifier for Multi-dimensional Distributions



- $\sum_1 = \sum_2$
- Linear Classifier
- Compare with means

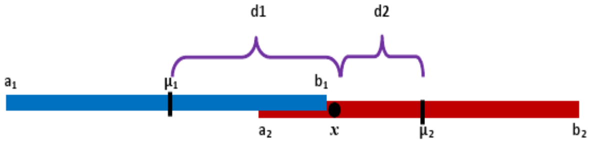
Quadratic Classifier for Multi-dimensional Distributions



- $\sum_1 \neq \sum_2$
- Quadratic Classifier
- Compare with means
- *Mahalanobis* distance

Importance of Mean

- Mean is the most central point of a distribution
- “Crucial Concept” of the Bayesian paradigm
- Yields the optimal accuracy



Bayesian Classifier

- Entire training set is required
- Learn μ_i, $\sum_i$ etc.
- Learn classifier

What to do in real life?

- Can we avoid “all” training data?
- Non-parametric - NN methods etc.

Bayesian Classifier

- Entire training set is required
- Learn μ_i, $\sum_i$ etc.
- Learn classifier

What to do in real life?

- Can we avoid “all” training data?
- Non-parametric - NN methods etc.

Bayesian Classifier

- Entire training set is required
- Learn μ_i, $\sum_i$ etc.
- Learn classifier

What to do in real life?

- Can we avoid “all” training data?
- Non-parametric - NN methods etc.

Prototype Reduction Schemes(PRSs)

- Reduce No. of training vectors - More than just mean
- Classifier built on the reduced set
- Performs nearly as well as that of the original set
- “Zillions” of PRSs (to quote Bezdek *et al*)
  - Condensed Nearest Neighbor (CNN)
  - Reduced Nearest Neighbor (RNN)
  - Prototypes for Nearest Neighbor (PNN)
  - Selective Nearest Neighbor (SNN)
  - Hybridized Kim-Oommen algorithm

Prototypes and Border Patterns

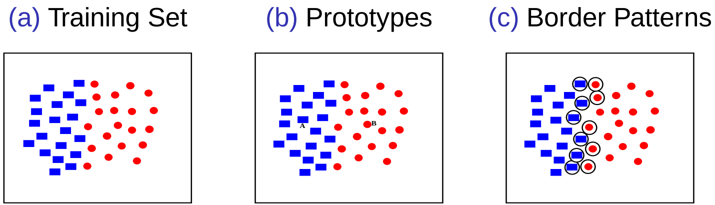


Figure: Border patterns vs Prototypes

- Border points above required in classification!
- Theoretical analysis: Undone and Unknown!

Difference between a Prototype and a Border Pattern

Attribute	Prototype	Border Pattern
Location	Anywhere in the training set	Near to the border
Information	Can represent the entire training set	Need not represent the entire set
Accuracy	Can attain near-optimal accuracy	Far borders are also needed

Table: Comparison between Prototypes and Border Patterns

Traditional BI Algorithms

- A subset of PRS
- Aims to obtain a Reference set containing Border points
  - Duch’s approach
  - Foody’s approach
  - Li’s approach

Drawbacks of Traditional BI Algorithms

- Traditional BI Algorithms obtain Near borders
- Not enough for the classification
- Needs Far Borders also
- Foody’s approach attains near-optimal accuracy
- But requires a large set of border patterns
    - This includes near and far borders
- No formal analysis!

Classification by Quantile Statistics

CQS

Basic Idea

- Given $p(\mathbf{x})$. What information is available?
  - Mean, Variance etc. - Moments/Central Moments
  - Used for traditional classification
- Is other information available?
  - Quantiles
  - Order Statistics
  - Unused in PR - Related to BI schemes!
  - We use *these* here

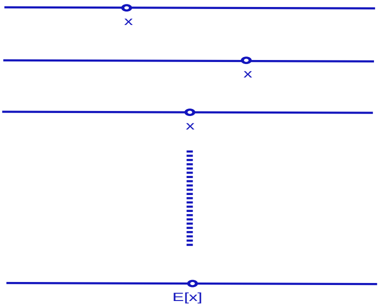
Basic Idea

- Given $p(\mathbf{x})$. What information is available?
  - Mean, Variance etc. - Moments/Central Moments
  - Used for traditional classification
- Is other information available?
  - Quantiles
  - Order Statistics
  - Unused in PR - Related to BI schemes!
  - We use *these* here

Basic Idea

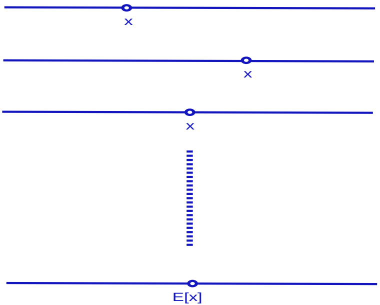
- Given $p(\mathbf{x})$. What information is available?
  - Mean, Variance etc. - Moments/Central Moments
  - Used for traditional classification
- Is other information available?
  - Quantiles
  - Order Statistics
  - Unused in PR - Related to BI schemes!
  - We use *these* here

Order Statistics: 1-OS



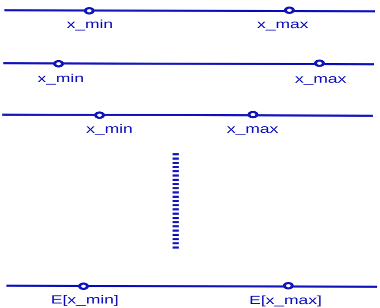
- 1-OS
- $E[\Phi(\mathbf{x}_{1,1})]=\frac{1}{2}$
- $E[x] = Mean$
- Leads to Bayesian Classifier

Order Statistics: 1-OS



- 1-OS
- $E[\Phi(\mathbf{x}_{1,1})] = \frac{1}{2}$
- $E[x] = Mean$
- Leads to Bayesian Classifier

Order Statistics: 2-OS



- 2-OS for U[0,1]
- $E[\Phi(\mathbf{x}_{1,2})] = \frac{1}{3} \implies E[\mathbf{x}_{1,2}] = \Phi^{-1}\left(\frac{1}{3}\right)$
- $E[\Phi(\mathbf{x}_{2,2})] = \frac{2}{3} \implies E[\mathbf{x}_{2,2}] = \Phi^{-1}\left(\frac{2}{3}\right)$
- *PR* Using 2-Order Stats??

Order Statistics - Fundamental Result

Consider a Random Variable **x** with pdf f and CDF F.

$$E[\mathbf{x}] = \int_{-\infty}^{+\infty} x \, f_{\mathbf{x}}(x) \, dx$$

$$E[\mathbf{x^k}] = \int_{-\infty}^{+\infty} x^k \, f_{\mathbf{x}}(x) \, dx$$

Order Statistics - Fundamental Result - Continued: 2

Now we want $E[\mathbf{x}_{r,n}]$.

$\mathbf{x}_{r,n}$ is the r^{th} OS out of n.

Question 1: What is $f_{\mathbf{x}}(x)$?

Question 2: What is $E[\mathbf{x}_{r,n}]$?

Order Statistics - Fundamental Result - Continued: 2

Now we want $E[\mathbf{x}_{r,n}]$.

$\mathbf{x}_{r,n}$ is the r^{th} OS out of n.

Question 1: What is $f_{\mathbf{x}}(x)$?

Question 2: What is $E[\mathbf{x}_{r,n}]$?

Order Statistics - Fundamental Result - Continued: 3

Let r^{th} OS be at $\mathbf{y} = \mathbf{x}_{r,n}$.

n, r are integers, $n \geq r \geq 2$.

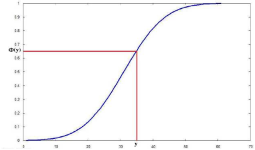
Φ is a nondecreasing and right-continuous function from $\mathbb{R} \rightarrow \mathbb{R}$.

Then, $E[\mathbf{x}_{r,n}]$ is

$$\frac{n!}{(r-1)!(n-r)!} \int_{-\infty}^{+\infty} y \underbrace{\Phi(y)^{r-1}}_{\text{(r-1) points}} \underbrace{(1-\Phi(y))^{n-r}}_{\text{(n-r) points}} \underbrace{\varphi(y)}_{f_{\mathbf{y}}(y)} dy$$

OS: Fundamental Result - Continued: 4

Why is this true?



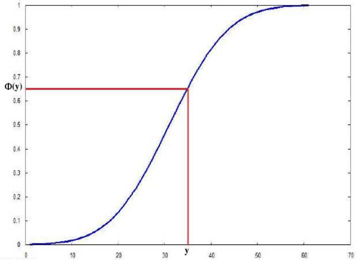
The result implies:

- $r-1$ smaller elements are drawn independently with prob. $\Phi(y)$.
- Other $n-r$ samples are drawn with probability $1-\Phi(y)$.
- Factorial terms: $(r-1)$ elements can be independently chosen from n elements.

OS: Fundamental Result - Continued: 5

Actually, the k^{th} moment is:

$$E[\mathbf{x}_{r,n}^k] = \frac{n!}{(r-1)!(n-r)!} \int_{-\infty}^{+\infty} y^k \Phi(y)^{r-1} (1-\Phi(y))^{n-r} \varphi(y) dy$$



OS: Fundamental Result - Continued: 6

The factorial terms and the integrals lead to:

$$E\left[\mathbf{x}_{r,n}\right] = \frac{B(r+1, n-r+1)}{B(r, n-r+1)} = \frac{r}{n+1}$$

and

$$E\left[\mathbf{x}_{r,n}^{k}\right] = \frac{B(r+k, n-r+1)}{B(r, n-r+1)} = \frac{n!\,(r+k-1)!}{(n+k)!\,(r-1)!}$$

CQS: Fundamental Result

Unfortunately, the inverses are not true:

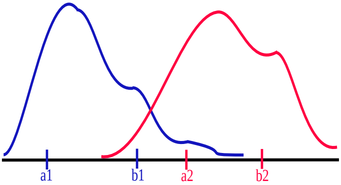
$$F^{-1}\left[\frac{r}{n+1}\right] \neq \mathbf{x}_{r,n}$$

and

$$F^{-1}\left[\frac{n!\,(r+k-1)!}{(n+k)!\,(r-1)!}\right] \neq \mathbf{x}_{r,n}^{k}$$

So: Move away from OS to Quantiles...

Generic Classifier



- a_1, b_1: QS for ω_1
- a_2, b_2: QS for ω_2

When test sample x^* comes:

- If $\mathbf{x}^* < \mathbf{b_1}$, $x^* \in \omega_1$
- If $\mathbf{x}^* > \mathbf{a_2}$, $x^* \in \omega_2$
- If $\mathbf{b_1} < \mathbf{x}^* < \mathbf{a_2}$, decision is based on $\|x^* - \mathbf{b_1}\|$ and $\|x^* - \mathbf{a_2}\|$

Symmetric Distributions

- Uniform Distribution
- Doubly-Exponential Distribution
- Gaussian Distribution
- Beta Distribution ($\alpha = \beta$)

Uniform Distribution

Pdfs of $\omega_1 \sim U[0,1]$ and $\omega_2 \sim U[h, 1+h]$ are:

$$f_1(x) = \begin{cases} \frac{1}{b_1 - a_1} & \text{if } a_1 \leq x \leq b_1; \\ 0 & \text{if } x < a_1 \text{ or } x > b_1 \end{cases}$$

$$f_2(x) = \begin{cases} \frac{1}{b_2 - a_2} & \text{if } a_2 \leq x \leq b_2; \\ 0 & \text{if } x < a_2 \text{ or } x > b_2 \end{cases}$$

Mean of ω_1 (i.e., μ_1): $\frac{1}{2}$

Mean of ω_2 (i.e., μ_2): $h + \frac{1}{2}$

2-QS CQS Points

For uniform distribution,

$$E\left[\mathbf{x}_{k,n}\right] = \frac{k}{n+1}$$

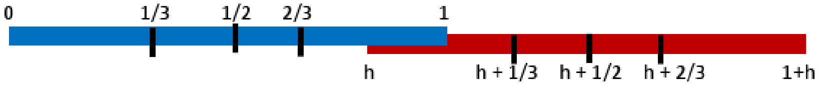
For ω_1, $E\left[\mathbf{x}_{1,2}\right] = \frac{1}{3}$ $E\left[\mathbf{x}_{2,2}\right] = \frac{2}{3}$

For ω_2, $E\left[\mathbf{x}_{1,2}\right] = h + \frac{1}{3}$ $E\left[\mathbf{x}_{2,2}\right] = h + \frac{2}{3}$

Theoretical Analysis: 2-QS

Bayes’	**CQS**
$D(x,\ \mu_1) < D(x,\ \mu_2)$	$D(x, q_1) < D(x, q_2)$
$\iff x - \frac{1}{2} < h + \frac{1}{2} - x$	$\iff D\left(x,\ \frac{2}{3}\right) < D\left(x,\ h + \frac{1}{3}\right)$
$\iff 2x < 1 + h$	$\iff x - \frac{2}{3} < h + \frac{1}{3} - x$
$\iff x < \frac{1+h}{2}$	$\iff x < \frac{1+h}{2}$

Pictorial Representation: 2-QS



Classifier

- $x < h \implies x \in \omega_1$
- $x > 1 \implies x \in \omega_2$
- $h < x < 1 \implies$ Test w.r.t $\frac{2}{3}$ and $h + \frac{1}{3}$

Experimental Results: 2-QS

h	0.95	0.90	0.85	0.80	0.75	0.70
Bayesian	97.58	95.1	92.42	90.23	87.82	85.4
CQS	97.58	95.1	92.42	90.23	87.82	85.4

Table: Classification of Uniformly Distributed Classes by the CQS 2-QS Method.

CQS points - k-QS

CQS achieves Bayes bound using symmetric pairs of the n-QS

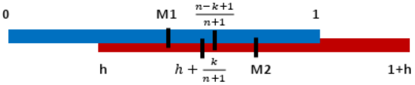
if and only if

$$k > \frac{(n+1)(1-h)}{2}.$$

$(n-k)$-QS for ω_1: $\frac{n-k+1}{n+1}$

k-QS for ω_2: $h + \frac{k}{n+1}$

Dual Condition - CQS



Theoretical Analysis: k-QS

Bayes'

$$D(x,\ \mu_1) \ < \ D(x,\ \mu_2)$$

$$\iff x-\tfrac{1}{2} \ < \ h+\tfrac{1}{2}-x$$

$$\iff 2x \ < \ 1+h$$

$$\iff x \ < \ \tfrac{1+h}{2}$$

CQS

$$D(x, q_1) < D(x, q_2)$$

$$\iff D\Big(x,\ \tfrac{n-k+1}{n+1}\Big) \ < \ D\Big(x,\ h+\tfrac{k}{n+1}\Big)$$

$$\iff x-\tfrac{n-k+1}{n+1} \ < \ h+\tfrac{k}{n+1}-x$$

$$\iff x \ < \ \tfrac{h+1}{2}$$

Experimental Results: k-QS

Trial No.	**Order(n)**	**Moments**	QS_1	QS_2	**CQS**	**CQS/ Dual CQS**
1	Two	$\{\frac{i}{3} \mid 1 \leq i \leq 2\}$	$\frac{2}{3}$	$h + \frac{1}{3}$	90.23	CQS
2	Three	$\{\frac{i}{4} \mid 1 \leq i \leq 3\}$	$\frac{3}{4}$	$h + \frac{1}{4}$	90.23	CQS
3	Four	$\{\frac{i}{5} \mid 1 \leq i \leq 4\}$	$\frac{4}{5}$	h + $\frac{1}{5}$	90.23	CQS
4	Five	$\{\frac{i}{6} \mid 1 \leq i \leq 5\}$	$\frac{4}{6}$	$h + \frac{2}{6}$	90.23	CQS
5	Six	$\{\frac{i}{7} \mid 1 \leq i \leq 6\}$	$\frac{4}{7}$	$h + \frac{2}{7}$	90.23	CQS
6	Nine	$\{\frac{i}{10} \mid 1 \leq i \leq 9\}$	$\frac{7}{10}$	$h + \frac{3}{10}$	90.23	CQS
7	Ten	$\{\frac{i}{11} \mid 1 \leq i \leq 10\}$	$\frac{10}{11}$	$h + \frac{1}{11}$	90.23	Dual CQS
8	Ten	$\{\frac{i}{11} \mid 1 \leq i \leq 10\}$	$\frac{9}{11}$	$h + \frac{2}{11}$	90.23	CQS
9	Ten	$\{\frac{i}{11} \mid 1 \leq i \leq 10\}$	$\frac{7}{11}$	$h + \frac{4}{11}$	90.23	CQS
10	Ten	$\{\frac{i}{11} \mid 1 \leq i \leq 10\}$	$\frac{6}{11}$	$h + \frac{5}{11}$	90.23	CQS

Table: Classification Obtained by using the Symmetric Pairs of the k-QS.

Dual Condition - CQS

For 10-QS in Table 3 where $h = 0.8$, QS used were

$$q_1 = \frac{10}{11} \text{ and } q_2 = h + \frac{1}{11}$$

Here, $1 - \frac{2k}{n+1} = 1 - \frac{2}{11} = 0.8182$

In this case, $\mathbf{h \not> 1 - \frac{2k}{n+1}}$.

Doubly Exponential Distribution

If pdfs of ω_1 and ω_2 are Doubly Exponentially distributed,

$$f_1(x) = \frac{\lambda_1}{2} e^{-\lambda_1 |x - c_1|}, \quad -\infty < x < \infty$$

$$f_2(x) = \frac{\lambda_2}{2} e^{-\lambda_2 |x - c_2|}, \quad -\infty < x < \infty$$

Mean of ω_1: c_1 Mean of ω_2: c_2

Variance of ω_1: $\frac{2}{\lambda_1^2}$ Variance of ω_2: $\frac{2}{\lambda_2^2}$

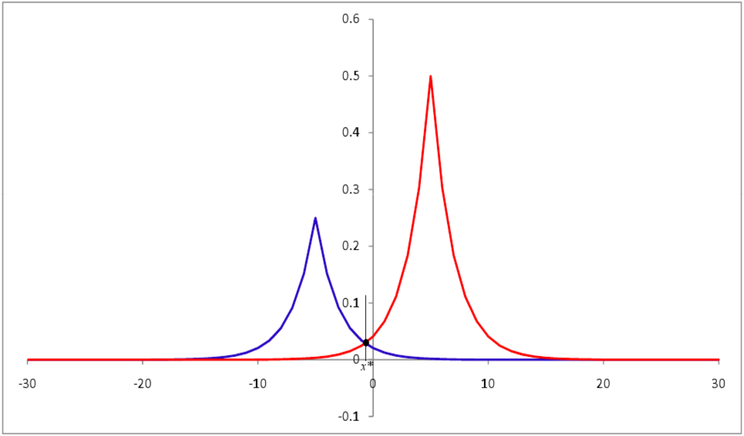


Figure: Doubly Exponential Distributions with different values for λ.

2-Order Statistics

q_1: $\frac{2}{3}^{rd}$ percentile of the first distribution
q_2: $\frac{1}{3}^{rd}$ percentile of the second distribution

q_1 and q_2: (Obtained by straightforward integrations etc.):

$$\begin{aligned} q_1 &= c_1 - \frac{1}{\lambda_1} log\left(\frac{2}{3}\right) \\ q_2 &= c_2 + \frac{1}{\lambda_2} log\left(\frac{2}{3}\right) \end{aligned}$$

Theoretical Analysis: 2-QS

Bayes'

$$p(x|\omega_1)P(\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(x|\omega_2)P(\omega_2)$$

$$\Longrightarrow \frac{\lambda}{2}\, e^{-\lambda|x-c_1|} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{\lambda}{2}\, e^{-\lambda|x-c_2|}$$

$$\Longrightarrow \lambda(x-c_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \lambda(c_2-x)$$

$$\Longrightarrow x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{c_1+c_2}{2}$$

CQS

$$D(x,q_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(x,q_2)$$

$$\Longrightarrow D\left(x,\ c_1-\frac{1}{\lambda}\ log\left(\frac{2}{3}\right)\right) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D\left(x,\ c_2+\frac{1}{\lambda}\ log\left(\frac{2}{3}\right)\right)$$

$$\Longrightarrow 2x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} c_1+c_2$$

$$\Longrightarrow x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{c_1+c_2}{2}$$

Experimental Results: - 2-QS

c_1	c_2	Bayes’	CQS
0	10	99.75	99.75
0	9	99.65	99.65
0	8	99.25	99.25
0	7	99.05	99.05
0	6	98.9	98.9
0	5	97.85	97.85
0	4	96.8	96.8
0	3	94.05	94.05
0	2	89.9	89.9

Table: Classification of Doubly Exponential Distribution by CQS Method.

CQS Points - k-QS

CQS achieves Bayes bound using symmetric pairs of the n-QS

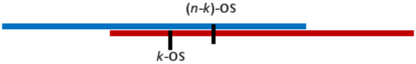
if and only if

$$log\left(\frac{2k}{n+1}\right) > \frac{c_1 - c_2}{2}.$$

$(n-k)$-QS for ω_1: $c_1 - log\left(\frac{2k}{n+1}\right)$

k-QS for ω_2: $c_2 + log\left(\frac{2k}{n+1}\right)$

Dual Condition - CQS



Theoretical Analysis: *k*-QS

Bayes'

$$p(x|\omega_1)P(\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(x|\omega_2)P(\omega_2)$$

$$\frac{\lambda}{2}e^{-\lambda(x-c_1)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{\lambda}{2}e^{-\lambda(x-c_2)}$$

$$\Longrightarrow x-c_1 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} x-c_2$$

$$\Longrightarrow x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{c_1+c_2}{2}$$

CQS

$$D(x,q_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(x,q_2)$$

$$\Longrightarrow D\left(x,c_1-log\left(\frac{2k}{n+1}\right)\right) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D\left(x,c_2+log\left(\frac{2k}{n+1}\right)\right)$$

$$\Longrightarrow x-\left(c_1-log\left(\frac{2k}{n+1}\right)\right) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \left(c_2-log\left(\frac{2k}{n+1}\right)\right)-x$$

$$\Longrightarrow x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{c_1+c_2}{2}$$

Experimental Results: k-QS

Order(n)	**Moments**	QS_1	QS_2	**CQS**	**CQS/ Dual CQS**
Two	$\left(\frac{2}{3}, \frac{1}{3}\right)$	$c_1 - \frac{1}{\lambda_1}\log\left(\frac{2}{3}\right)$	$c_2 + \frac{1}{\lambda_2}\log\left(\frac{2}{3}\right)$	95.2	CQS
Three	$\left(\frac{3}{4}, \frac{1}{4}\right)$	$c_1 - \frac{1}{\lambda_1}\log\left(\frac{1}{2}\right)$	$c_2 + \frac{1}{\lambda_2}\log\left(\frac{1}{2}\right)$	95.2	CQS
Four	$\left(\frac{5-i}{5}, \frac{i}{5}\right), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1}\log\left(\frac{4}{5}\right)$	$c_2 + \frac{1}{\lambda_2}\log\left(\frac{4}{5}\right)$	95.2	CQS
Five	$\left(\frac{6-i}{6}, \frac{i}{6}\right), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1}\log\left(\frac{1}{3}\right)$	$c_2 + \frac{1}{\lambda_2}\log\left(\frac{1}{3}\right)$	95.2	CQS
Six	$\left(\frac{7-i}{7}, \frac{i}{7}\right) 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1}\log\left(\frac{4}{7}\right)$	$c_2 + \frac{1}{\lambda_2}\log\left(\frac{4}{7}\right)$	95.2	CQS
Seven	$\left(\frac{8-i}{8}, \frac{i}{8}\right) 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1}\log\left(\frac{1}{4}\right)$	$c_2 + \frac{1}{\lambda_2}\log\left(\frac{1}{4}\right)$	95.2	CQS
Eight	$\left(\frac{9-i}{9}, \frac{i}{9}\right), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1}\log\left(\frac{2}{9}\right)$	$c_2 + \frac{1}{\lambda_2}\log\left(\frac{2}{9}\right)$	95.2	Dual CQS
Eight	$\left(\frac{9-i}{9}, \frac{i}{9}\right), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1}\log\left(\frac{4}{9}\right)$	$c_2 + \frac{1}{\lambda_2}\log\left(\frac{4}{9}\right)$	95.2	CQS
Nine	$\left(\frac{10-i}{10}, \frac{i}{10}\right), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1}\log\left(\frac{3}{5}\right)$	$c_2 + \frac{1}{\lambda_2}\log\left(\frac{3}{5}\right)$	95.2	CQS

Table: Classification using symmetric QS pairs for different orders. ($c_1=0, c_2=3, 1 \leq i \leq \frac{n}{2}$)

Dual Condition - CQS

For 8-QS in Table 5, QS used were

$$c_1 - \frac{1}{\lambda_1} log\left(\frac{2}{9}\right) \text{ and } c_2 + \frac{1}{\lambda_2} log\left(\frac{2}{9}\right)$$

Here, $log\left(\frac{2k}{n+1}\right) = -1.504$ and $\frac{c_1 - c_2}{2} = -1.5$

In this case, $log\left(\frac{2k}{n+1}\right) \not> \frac{c_1 - c_2}{2}$.

Gaussian Distribution

If pdfs of ω_1 and ω_2 are normally distributed,

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu_1)^2}{2\sigma^2}}$$

$$f_2(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu_2)^2}{2\sigma^2}}$$

Mean of ω_1: μ_1

Variance of ω_1: σ^2

Mean of ω_2: μ_2

Variance of ω_2: σ^2

2-Order Statistics

q_1: $\frac{2}{3}^{rd}$ percentile of the first distribution
q_2: $\frac{1}{3}^{rd}$ percentile of the second distribution

q_1 and q_2:

$$q_1 = \mu_1 + \frac{\sigma}{\sqrt{2\pi}}$$
$$q_2 = \mu_2 - \frac{\sigma}{\sqrt{2\pi}}$$

Theoretical Analysis: 2-QS

Bayes'

$$p(x|\omega_1)P(\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(x|\omega_2)P(\omega_2)$$

$$\Longrightarrow \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu_1)^2}{2\sigma^2}} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu_2)^2}{2\sigma^2}}$$

$$\Longrightarrow x-\mu_1 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \mu_2 - x$$

$$\Longrightarrow x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{\mu_1+\mu_2}{2}$$

CQS

$$D(x,\ q_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(x,\ q_2)$$

$$\Longrightarrow D\left(x,\ \mu_1-\frac{\sigma}{\sqrt{2\pi}}\right) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D\left(x,\ \mu_2+\frac{\sigma}{\sqrt{2\pi}}\right)$$

$$\Longrightarrow x-\left(\mu_1-\frac{\sigma}{\sqrt{2\pi}}\right) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \left(\mu_2+\frac{\sigma}{\sqrt{2\pi}}\right)-x$$

$$\Longrightarrow 2x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \mu_1+\mu_2$$

$$\Longrightarrow x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{\mu_1+\mu_2}{2}$$

Experimental Results: 2-QS

μ_1	0	0	0	0	0	0
μ_2	14	12	10	8	6	4
Bayes’	99.2	96.5	95.1	95	90	85
CQS	99.2	96.5	95.1	95	90	85

Table: Classification of Normally distributed classes by the CQS 2-QS method.

CQS points -k-QS

- Gaussian pdf is not integrable
- Closed form expression for the condition: Not available
- Solve numerically or by inverse function

k-QS CQS points by Inverse Function

$$g(u) = \sqrt{-2\ logu} \cdot \frac{A\sqrt{-2\ logu}}{B\sqrt{-2\ logu}},$$

where $A(x) = \sum_{i=0}^{4} a_i x^i$, $B(x) = \sum_{i=0}^{4} b_i x^i$, and where:

i	a_i	b_i
0	-0.3222232431088	0.0993484626060
1	-1.0	0.588581570495
2	-0.342242088547	0.531103462366
3	-0.0204231210245	0.103537752850
4	-0.0000453642210148	0.0038560700634

Table: Coefficients for the inverse Normal function.

Experimental Results: k-QS

No.	Order(n)	Moments	CQS	CQS/ Dual CQS
1	Two	$\left(\frac{2}{3}, \frac{1}{3}\right)$	91.865	CQS
2	Four	$\left(\frac{4}{5}, \frac{1}{5}\right)$	91.865	CQS
3	Six	$\left(\frac{6}{7}, \frac{1}{7}\right)$	91.865	CQS
4	Eight	$\left(\frac{8}{9}, \frac{1}{9}\right)$	91.865	CQS
5	Ten	$\left(\frac{10}{11}, \frac{1}{11}\right)$	91.865	Dual CQS
6	Ten	$\left(\frac{9}{11}, \frac{2}{11}\right)$	91.865	CQS
7	Twelve	$\left(\frac{12}{13}, \frac{1}{13}\right)$	91.865	Dual CQS
8	Twelve	$\left(\frac{10}{13}, \frac{3}{13}\right)$	91.865	CQS

Table: Results of the classification obtained by CQS - k-QS for various n.

Beta Distribution

No	α, β	Distribution
1	$\alpha = \mathbf{1},\ \beta = \mathbf{1}$	**Uniform Distribution**
2	$\alpha < 1,\ \beta < 1$	U-shaped
3	$\alpha = 1,\ \beta = 2$	Straight line
4	$\alpha = \beta$	**Symmetric about $\frac{1}{2}$**
5	$\alpha = \frac{1}{2},\ \beta = \frac{1}{2}$	Arcsine Distribution
6	$\alpha > 1,\ \beta > 1$	Unimodal Distribution
7	$\alpha = 1,\ \beta > 1$	Strictly Convex
8	$\alpha = 1,\ 1 < \beta < 2$	Strictly Concave

Table: Different forms of the Beta distribution for the various values of α and β.

Beta Distribution ($\alpha = \beta$)

If pdfs of ω_1 and ω_2 are Beta distributed, for simplicity, let them be:

$$
\begin{aligned}
f(x, 2, 2) &= 6x(1-x) \\
f(x-\theta, 2, 2) &= 6(x-\theta)(1-x+\theta)
\end{aligned}
$$

Experimental Results: 2-QS

θ	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
Bayesian	99.845	99.45	98.185	96.94	94.95	92.86	90.31	88.075
CQS	99.845	99.45	98.185	96.94	94.95	92.86	90.31	88.075

Table: Classification of Beta (Beta(2,2)) distributed classes by the CQS 2-QS method.

Experimental Results: k-QS

Trial No.	Order(n)	Moments	QS_1	QS_2	CQS	CQS/ Dual CQS
1	Two	$\langle \frac{2}{3}, \frac{1}{3} \rangle$	0.61304	$\theta + 0.38696$	87.3	CQS
2	Four	$\langle \frac{4}{5}, \frac{1}{5} \rangle$	0.71286	$\theta + 0.28714$	87.3	CQS
3	Four	$\langle \frac{3}{5}, \frac{2}{5} \rangle$	0.56707	$\theta + 0.43293$	87.3	CQS
4	Six	$\langle \frac{6}{7}, \frac{1}{7} \rangle$	0.7621	$\theta + 0.23790$	87.3	CQS
5	Six	$\langle \frac{5}{7}, \frac{2}{7} \rangle$	0.6471	$\theta + 0.3529$	87.3	CQS
6	Six	$\langle \frac{4}{7}, \frac{3}{7} \rangle$	0.54776	$\theta + 0.45224$	87.3	CQS
7	Eight	$\langle \frac{8}{9}, \frac{1}{9} \rangle$	0.79269	$\theta + 0.20731$	87.3	Dual CQS
8	Eight	$\langle \frac{7}{9}, \frac{2}{9} \rangle$	0.69508	$\theta + 0.30492$	87.3	CQS
9	Eight	$\langle \frac{5}{9}, \frac{4}{9} \rangle$	0.53711	$\theta + 0.46289$	87.3	CQS

Table: Results of the classification obtained by CQS - k-QS for various n.

Asymmetric Distributions

- Rayleigh Distribution
- Gamma Distribution
- Beta Distribution ($\alpha \neq \beta$)

Rayleigh Distribution

The pdf of Rayleigh distribution is: $f(x, \sigma) = \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}$.

Mean: $\sigma\sqrt{\frac{\pi}{2}}$ Variance: $\frac{4-\pi}{2}\sigma^2$

CQS positions q_1 and q_2:

$$\begin{aligned} q_1 &= \sigma\sqrt{2\,ln(3)} \\ q_2 &= \theta + \sigma\sqrt{2ln\left(\frac{3}{2}\right)} \end{aligned}$$

Theoretical Analysis: 2-QS

Bayes'

$$p(x|\omega_1)P(\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(x|\omega_2)P(\omega_2)$$

$$\Longrightarrow \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{x-\theta}{\sigma^2} e^{\frac{-(x-\theta)^2}{2\sigma^2}}$$

$$\Longrightarrow \frac{x}{x-\theta} \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} e^{\frac{-(x-\theta)^2}{2\sigma^2}+\frac{x^2}{2\sigma^2}}$$

$$\Longrightarrow ln\left(\frac{x}{x-\theta}\right) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{-\theta^2+2\theta x}{2\sigma^2}$$

CQS

$$D(x,q_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(x,q_2)$$

$$\Longrightarrow D\left(x,\ \sigma\sqrt{2\ ln(3)}\right) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D\left(x,\ \theta+\sigma\sqrt{2ln\left(\frac{3}{2}\right)}\right)$$

$$\Longrightarrow 2x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \theta+\sigma\sqrt{2ln(3)}+\sigma\sqrt{2ln\left(\frac{3}{2}\right)}$$

$$\Longrightarrow x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{\theta}{2}+\frac{\sigma}{\sqrt{2}}\left(\sqrt{ln(3)}+\sqrt{ln\left(\frac{3}{2}\right)}\right)$$

Rayleigh Distribution - Pictorial Representation

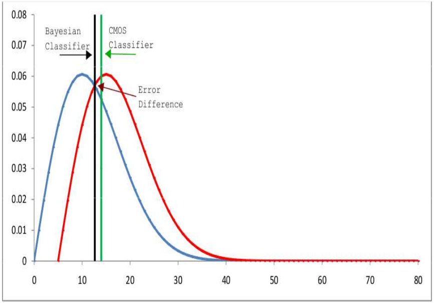


Figure: The differences of the error probabilities.

Rayleigh Distribution - Error Analysis

θ	1	1.5	2	2.5	3
a	3	4	5	5.3	6.5
Max. Bounded Error(in %)	0.15	0.06	0.05	0.001	0

Table: The maximum bounded error by the CQS classifier when compared to the Bayesian classifier.

Experimental Results: 2-QS

θ	3	2.5	2	1.5	1
Bayesian	99.1	97.35	94.45	87.75	78.80
CQS	99.1	97.35	94.40	87.70	78.65

Table: Classification of Rayleigh distributed classes by the CQS 2-QS method.

CQS Points - k-QS

k-QS CQS positions

$(n-k)$-QS for ω_1: $\sigma\sqrt{2\ ln\left(\frac{n+1}{k}\right)}$

k-QS for ω_2: $\theta + \sigma\sqrt{2\ ln\left(\frac{n+1}{n+1-k}\right)}$

CQS/Dual CQS achieves near-optimal Bayes bound

CQS **if** $\sqrt{ln\left(\frac{n+1}{k}\right)} - \sqrt{ln\left(\frac{n+1}{n+1-k}\right)} < \frac{\theta}{\sigma\sqrt{2}}$

Dual CQS **if** $\sqrt{ln\left(\frac{n+1}{k}\right)} - \sqrt{ln\left(\frac{n+1}{n+1-k}\right)} > \frac{\theta}{\sigma\sqrt{2}}$

Experimental Results: k-QS

Order(n)	**Moments**	QS_1	QS_2	**CQS**	**CQS/ Dual CQS**
Two	$\left(\frac{2}{3}, \frac{1}{3}\right)$	$\sigma\sqrt{\left(2\ ln\left(\frac{3}{1}\right)\right)}$	$\theta + \sigma\sqrt{\left(2\ ln\left(\frac{3}{2}\right)\right)}$	82.05	CQS
Four	$\left(\frac{5-i}{5}, \frac{i}{5}\right), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{5}{2}\right)\right)}$	$\theta + \sigma\sqrt{\left(2\ ln\left(\frac{5}{3}\right)\right)}$	82.0	CQS
Six	$\left(\frac{7-i}{7}, \frac{i}{7}\right) 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{7}{1}\right)\right)}$	$\theta + \sigma\sqrt{\left(2\ ln\left(\frac{7}{6}\right)\right)}$	81.6	Dual CQS
Six	$\left(\frac{7-i}{7}, \frac{i}{7}\right) 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{7}{2}\right)\right)}$	$\theta + \sigma\sqrt{\left(2\ ln\left(\frac{7}{5}\right)\right)}$	82.10	CQS
Six	$\left(\frac{7-i}{7}, \frac{i}{7}\right) 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{7}{3}\right)\right)}$	$\theta + \sigma\sqrt{\left(2\ ln\left(\frac{7}{4}\right)\right)}$	82.15	CQS
Eight	$\left(\frac{9-i}{9}, \frac{i}{9}\right), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{9}{1}\right)\right)}$	$\theta + \sigma\sqrt{\left(2\ ln\left(\frac{9}{8}\right)\right)}$	81.55	Dual CQS
Eight	$\left(\frac{9-i}{9}, \frac{i}{9}\right), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{9}{2}\right)\right)}$	$\theta + \sigma\sqrt{\left(2\ ln\left(\frac{9}{7}\right)\right)}$	82.05	CQS

Table: Results of the classification obtained by CQS - k-QS for various n.

Gamma Distribution

If pdfs of ω_1 and ω_2 are Gamma distributed, for simplicity, let them be:

$$f(x,2,1) = x\ e^{-x}$$

$$f(x-\theta,2,1) = (x-\theta)\ e^{-(x-\theta)}$$

CQS positions obtained using Octave:

$q_1 = 2.2893$ $\quad\quad$ $q_2 = \theta + 1.1888$

Theoretical Analysis: 2-QS

Median

$$D(x,\nu_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(x,\nu_2)$$

$$\implies D(x,1.6783) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(x,1.6783+\theta)$$

$$\implies x-1.6783 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 1.6783+\theta-x$$

$$\implies x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 1.6783+\frac{\theta}{2}$$

CQS

$$D(x,q_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(x,q_2)$$

$$\implies D(x,2.2893) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(x,1.1888+\theta)$$

$$\implies x-2.2893 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 1.1888+\theta-x$$

$$\implies x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 1.7391+\frac{\theta}{2}$$

Experimental Results: 2-QS

n	Median	CQS
4.5	94.825	95.01
4.0	94.25	94.49
3.5	92.74	92.915
3.0	90.765	90.425
2.5	86.51	85.985
2.0	80.145	79.54
1.5	72.64	72.34

Table: Classification of Gamma distributed classes by the CQS 2-QS method.

Experimental Results: k-QS

Classifier	**Moments**	$\theta =$ **4.5**	**4.0**	**3.5**	**3.0**	**2.5**	**2.0**
Bayes	-	97.06	95.085	93.145	90.68	86.93	81.53
Mean	-	96.165	94.875	92.52	88.335	83.105	77.035
Median	-	96.04	93.57	92.735	90.775	86.275	80.115
2-QS	$\left(\frac{2}{3}, \frac{1}{3}\right)$	95.285	93.865	92.87	90.61	86.085	79.48
4-QS	$\left(\frac{4}{5}, \frac{1}{5}\right)$	95.905	94.605	93.11	89.57	84.68	77.875 (D)
4-QS	$\left(\frac{3}{5}, \frac{2}{5}\right)$	95.185	93.675	92.82	**90.855**	86.02	**80.32**
6-QS	$\left(\frac{6}{7}, \frac{1}{7}\right)$	96.405	**95.01**	92.125	88.005	82.71 (D)	76.435 (D)
6-QS	$\left(\frac{5}{7}, \frac{2}{7}\right)$	95.47	94.11	**93.135**	90.16	85.495	79.55
6-QS	$\left(\frac{4}{7}, \frac{3}{7}\right)$	95.135	93.625	92.78	90.745	**86.135**	80.165
8-QS	$\left(\frac{8}{9}, \frac{1}{9}\right)$	**96.815**	94.895	91.555	86.905 (D)	80.59 (D)	75.94 (D)
8-QS	$\left(\frac{7}{9}, \frac{2}{9}\right)$	95.8	94.445	93.11	89.885	84.81	78.535
8-QS	$\left(\frac{5}{9}, \frac{4}{9}\right)$	95.135	93.625	92.735	90.7	86.085	80.045

Table: Results of the classification obtained by CQS - k-QS for various n.

Beta Distribution

No	α, β	Distribution
1	$\alpha = 1, \ \beta = 1$	Uniform Distribution
2	$\alpha < 1, \ \beta < 1$	U-shaped
3	$\alpha = 1, \ \beta = 2$	Straight line
4	$\alpha = \beta$	Symmetric about $\frac{1}{2}$
5	$\alpha = \frac{1}{2}, \ \beta = \frac{1}{2}$	Arcsine Distribution
6	$\alpha > 1, \ \beta > 1$	**Unimodal Distribution**
7	$\alpha = 1, \ \beta > 1$	Strictly Convex
8	$\alpha = 1, \ 1 < \beta < 2$	Strictly Concave

Table: Different forms of the Beta distribution for the various values of α and β.

Beta Distribution ($\alpha > 1, \beta > 1$)

If pdfs of ω_1 and ω_2 are Beta distributed, for simplicity, let them be:
(when $\alpha = 2$ and $\beta = 5$):

$$f(x, 2, 5) = 30x(1 - x)^4$$

$$f(x - \theta, 2, 5) = 30(x - \theta)(1 - x + \theta)^4$$

CQS positions obtained using Octave:

$$q_1 = 0.34249 \qquad q_2 = \theta + 0.1954$$

Medians obtained using Octave:

$$\nu_1 = 0.26445 \qquad \nu_2 = \theta + 0.26445$$

Experimental Results: 2-QS

θ	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8
Median	89.625	92.9	94.3	95.525	97.3	97.975	98.375	99.05	99.15
CQS	89.475	92.775	94.525	95.75	97.3	98.05	98.375	99.2	99.225

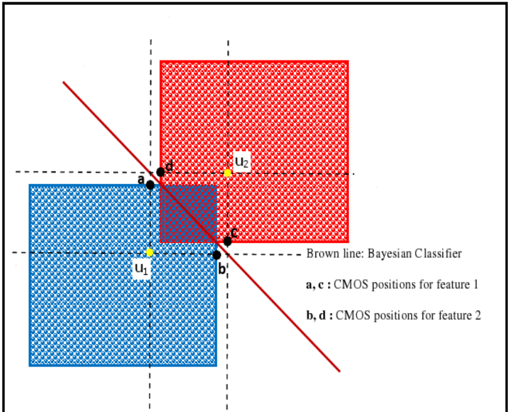
Table: Classification of Beta (Beta(2,5)) distributed classes by the CQS 2-QS method.

Experimental Results: k-QS

Classifier	**Moments**	$\theta =$ **0.35**	**0.45**	**0.55**	**0.65**	**0.75**	**0.85**
Mean	-	85.325	92.575	96.55	98.3	99.4	99.475
Median	-	86.675	92.775	95.525	97.975	99.05	99.275
2-QS	$\left(\frac{2}{3}, \frac{1}{3}\right)$	86.2	92.575	95.75	98.05	99.2	99.275
4-QS	$\left(\frac{4}{5}, \frac{1}{5}\right)$	85.375	92.525	96.225	98.225	99.325	99.475
4-QS	$\left(\frac{3}{5}, \frac{2}{5}\right)$	86.475	**92.775**	95.6	98.05	99.125	99.275
6-QS	$\left(\frac{6}{7}, \frac{1}{7}\right)$	85.2 (D)	92.425	**96.475**	**98.35**	99.45	99.625
6-QS	$\left(\frac{5}{7}, \frac{2}{7}\right)$	86.125	92.625	96.0	98.075	99.2	99.275
6-QS	$\left(\frac{4}{7}, \frac{3}{7}\right)$	86.55	**92.775**	95.525	97.975	99.125	**99.75**
8-QS	$\left(\frac{8}{9}, \frac{1}{9}\right)$	84.225 (D)	92.225	96.225	98.35	**99.5**	99.375
8-QS	$\left(\frac{7}{9}, \frac{2}{9}\right)$	85.675	92.5	96.175	98.15	99.325	99.375
8-QS	$\left(\frac{5}{9}, \frac{4}{9}\right)$	**86.575**	**92.775**	95.525	97.975	99.125	99.275

Table: Results of the classification obtained by CQS - k-QS for various n.

Two-dimensional Uniform Distribution



2-QS CQS Points

For 2-D Uniform distribution in $[0,1]^2$ and $[h, 1+h]^2$,

2-QS CQS Points are:

$q_1 = \left(\frac{2}{3}, \frac{2}{3}\right) \qquad q_2 = \left(h + \frac{1}{3}, h + \frac{1}{3}\right)$

whereas Means are:

$M_1 = \left(\frac{1}{2}, \frac{1}{2}\right) \qquad M_2 = \left(h + \frac{1}{2}, h + \frac{1}{2}\right)$

Theoretical Analysis: 2-QS

Bayes'

$$D(X,M_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(X,M_2)$$

$$\Longleftrightarrow \sqrt{\left(x_1-\tfrac{1}{2}\right)^2+\left(x_2-\tfrac{1}{2}\right)^2}$$

$$\underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \sqrt{\left(x_1-\left(h+\tfrac{1}{2}\right)\right)^2+\left(x_2-\left(h+\tfrac{1}{2}\right)\right)^2}$$

$$\Longleftrightarrow 2h\ (x_1+x_2) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 2h\ (h+1)$$

$$\Longleftrightarrow x_1+x_2 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} h+1$$

CQS

$$D(X,q_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(X,q_2)$$

$$\Longleftrightarrow \sqrt{\left(x_1-\tfrac{2}{3}\right)^2+\left(x_2-\tfrac{2}{3}\right)^2}$$

$$\underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \sqrt{\left(x_1-\left(h+\tfrac{1}{3}\right)\right)^2+\left(x_2-\left(h+\tfrac{1}{3}\right)\right)^2}$$

$$\Longleftrightarrow 2(x_1+x_2)\left(h-\tfrac{1}{3}\right) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 2\left(h^2+\tfrac{2}{3}h-\tfrac{3}{9}\right)$$

$$\Longleftrightarrow x_1+x_2 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} h+1$$

2-D Uniform Distribution -Experimental Results

h	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
Bayesian	99.85	99.51	98.88	98.05	97.15	95.56	94.14	91.82
CQS	99.85	99.51	98.88	98.05	97.15	95.56	94.14	91.82

Table: Classification of Uniformly Distributed 2-D Classes by the CQS 2-QS.

Theoretical Analysis: k-QS

Bayes'

$$D(X,M_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(X,M_2)$$

$$\Longleftrightarrow \sqrt{\left(x_1-\tfrac{1}{2}\right)^2+\left(x_2-\tfrac{1}{2}\right)^2} \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \sqrt{\left(x_1-\left(h+\tfrac{1}{2}\right)\right)^2+\left(x_2-\left(h+\tfrac{1}{2}\right)\right)^2}$$

$$\Longleftrightarrow 2h\ (x_1+x_2) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 2h\ (h+1)$$

$$\Longleftrightarrow x_1+x_2 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} h+1$$

CQS

$$D(X,q_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(X,q_2)$$

$$\Longleftrightarrow \sqrt{\left(x_1-\tfrac{n+1-k}{n+1}\right)^2+\left(x_2-\tfrac{n+1-k}{n+1}\right)^2} \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \sqrt{\left(x_1-\left(h+\tfrac{k}{n+1}\right)\right)^2+\left(x_2-\left(h+\tfrac{k}{n+1}\right)\right)^2}$$

$$\Longleftrightarrow 2(x_1+x_2)\ \left(h-1+\tfrac{2k}{n+1}\right) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 2\left(h-1+\tfrac{2k}{n+1}\right)\ (h+1)$$

$$\Longleftrightarrow x_1+x_2 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} h+1$$

2-D Uniform Distribution -Experimental Results

$h \longrightarrow$	**0.95**	**0.90**	**0.85**	**0.80**	**0.75**	**0.70**	**0.65**
$\langle \frac{2}{3}, \frac{1}{3} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69	93.73
$\langle \frac{4}{5}, \frac{1}{5} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69	93.73
$\langle \frac{6}{7}, \frac{1}{7} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69 (D)	93.73 (D)
$\langle \frac{5}{7}, \frac{2}{7} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69	93.73
$\langle \frac{4}{7}, \frac{3}{7} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69	93.73
$\langle \frac{8}{9}, \frac{1}{9} \rangle$	99.92	99.58	98.86	97.94	96.78 (D)	95.69 (D)	93.73 (D)
$\langle \frac{7}{9}, \frac{2}{9} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69	93.73
$\langle \frac{5}{9}, \frac{4}{9} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69	93.73

Table: Classification of Uniformly Distributed 2-D Classes by the CQS k-QS.

Multi-dimensional Uniform Distribution

For the d-dimensional Uniform distributions in $U[0,1]^d$ and $U[h, 1+h]^d$, the classifier is:

$$x_1 + x_2 + \ldots + x_d \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{d}{2}(h+1).$$

CQS/Dual CQS attains Bayes' bound.

2-D Doubly-Exponential Distribution

$$f_1(X) = \frac{\lambda_{11}}{2} e^{-\lambda_{11}|x_1 - \mathbf{c_{11}}|} \cdot \frac{\lambda_{12}}{2} e^{-\lambda_{12}|x_2 - \mathbf{c_{12}}|}, \quad -\infty < x_1 < \infty, -\infty < x_2 < \infty$$

$$f_2(X) = \frac{\lambda_{21}}{2} e^{-\lambda_{21}|x_1 - \mathbf{c_{21}}|} \cdot \frac{\lambda_{22}}{2} e^{-\lambda_{22}|x_2 - \mathbf{c_{22}}|}, \quad -\infty < x_1 < \infty, -\infty < x_2 < \infty$$

Means:

$C_1 = (c_{11}, c_{12}) \qquad C_2 = (c_{21}, c_{22})$

CQS Positions:

$$q_1 = \left[c_{11} - \frac{1}{\lambda_{11}} \, ln\left(\frac{2k}{n+1}\right), \; c_{12} - \frac{1}{\lambda_{12}} \, ln\left(\frac{2k}{n+1}\right)\right]^T$$

$$q_2 = \left[c_{21} + \frac{1}{\lambda_{21}} \, ln\left(\frac{2k}{n+1}\right), \; c_{22} + \frac{1}{\lambda_{22}} \, ln\left(\frac{2k}{n+1}\right)\right]^T$$

Theoretical Analysis: k-QS

Bayes'

$$p(X|\omega_1)\ P(\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(X|\omega_2)\ P(\omega_2)$$

$$\Longrightarrow \frac{\lambda}{2}\ e^{-\lambda|x_1-c_{11}|}.\frac{\lambda}{2}\ e^{-\lambda|x_2-c_{12}|}$$

$$\underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{\lambda}{2}\ e^{-\lambda|x_1-c_{21}|}.\frac{\lambda}{2}\ e^{-\lambda|x_2-c_{22}|}$$

$$\Longrightarrow x_1+x_2-c_{21}-c_{22} \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} -x_1-x_2+c_{11}+c_{12}$$

$$\Longrightarrow x_1+x_2 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} c$$

CQS

$$D(X,q_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(X,q_2)$$

$$\Longrightarrow \sqrt{\left(x_1+ln\left(\frac{2k}{n+1}\right)\right)^2+\left(x_2+ln\left(\frac{2k}{n+1}\right)\right)^2}$$

$$\underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \sqrt{\left(\left(c+ln\left(\frac{2k}{n+1}\right)\right)-x_1\right)^2+\left(\left(c+ln\left(\frac{2k}{n+1}\right)\right)-x_2\right)^2}$$

$$\Longrightarrow 2(x_1+x_2)\left(c+2ln\left(\frac{2k}{n+1}\right)\right) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 2c\left(c+2ln\left(\frac{2k}{n+1}\right)\right)$$

$$\Longrightarrow x_1+x_2 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} c$$

2-D Doubly-Exponential Distribution - Results

No.	c	w.r.t Mean	$\left\langle \frac{2}{3}, \frac{1}{3} \right\rangle$	$\left\langle \frac{4}{5}, \frac{1}{5} \right\rangle$	$\left\langle \frac{5}{7}, \frac{2}{7} \right\rangle$	$\left\langle \frac{8}{9}, \frac{1}{9} \right\rangle$
1	3	96.55	96.55	96.55	96.55	96.55
2	2.5	95.5	95.5	95.5	95.5	95.5
3	2	92	92	92	92	92
4	1.5	89.3	89.3	89.3 (D)	89.3	89.3 (D)

Table: Classification of Doubly-Exponentially Distributed 2-D Classes by the CQS 2-QS.

Multi-dimensional Doubly-Exponential Distribution

For the d-dimensional identical and symmetric Doubly-Exponential distributions, the classifier is:

$$x_1 + x_2 + ... + x_d = \frac{d}{2} \cdot c.$$

CQS/Dual CQS attains Bayes’ bound.

2-D Gaussian Distribution

$$f_1(X) = \frac{1}{\sqrt{2\pi}\sigma_{11}} e^{\frac{-(x_1-\mu_{11})^2}{2\sigma_{11}^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_{12}} e^{\frac{-(x_2-\mu_{12})^2}{2\sigma_{12}^2}}, \quad -\infty < x_1 < \infty, -\infty < x_2 < \infty$$

$$f_2(X) = \frac{1}{\sqrt{2\pi}\sigma_{21}} e^{\frac{-(x_1-\mu_{21})^2}{2\sigma_{21}^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_{22}} e^{\frac{-(x_2-\mu_{22})^2}{2\sigma_{22}^2}}, \quad -\infty < x_1 < \infty, -\infty < x_2 < \infty$$

Means:

$M_1 = (\mu_{11}, \mu_{12})$ $\quad$ $M_2 = (\mu_{21}, \mu_{22})$

CQS Positions:

$$q_1 = \left[\mu_{11} - \frac{\sigma_{11}}{\sqrt{2\pi}},\ \mu_{12} - \frac{\sigma_{12}}{\sqrt{2\pi}}\right]^T$$

$$q_2 = \left[\mu_{21} - \frac{\sigma_{21}}{\sqrt{2\pi}},\ \mu_{22} - \frac{\sigma_{22}}{\sqrt{2\pi}}\right]^T$$

Theoretical Analysis: k-QS

Bayes’

$$p(X|\omega_1)\ P(\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(X|\omega_2)\ P(\omega_2)$$

$$\Longrightarrow \frac{1}{\sqrt{2\pi}\sigma_{11}} e^{\frac{-(x_1-\mu_{11})^2}{2\sigma_{11}^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_{12}} e^{\frac{-(x_2-\mu_{12})^2}{2\sigma_{12}^2}}$$

$$\underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{1}{\sqrt{2\pi}\sigma_{21}} e^{\frac{-(x_1-\mu_{21})^2}{2\sigma_{21}^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_{22}} e^{\frac{-(x_2-\mu_{22})^2}{2\sigma_{22}^2}}$$

$$\Longrightarrow x_1^2+x_2^2 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} (x_1-\mu)^2+(x_2-\mu)^2$$

$$\Longrightarrow x_1+x_2 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \mu$$

CQS

$$D(X,q_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} D(X,q_2)$$

$$\Longrightarrow \sqrt{\left(x_1+\frac{\sigma}{\sqrt{2\pi}}\right)^2+\left(x_2+\frac{\sigma}{\sqrt{2\pi}}\right)^2}$$

$$\underset{\omega_2}{\overset{\omega_1}{\gtrless}} \sqrt{\left(\mu+\frac{\sigma}{\sqrt{2\pi}}-x_1\right)^2+\left(\mu+\frac{\sigma}{\sqrt{2\pi}}-x_2\right)^2}$$

$$\Longrightarrow 2(x_1+x_2)\left(\mu+\frac{2\sigma}{\sqrt{2\pi}}\right) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 2\mu\left(\mu+\frac{2\sigma}{\sqrt{2\pi}}\right)$$

$$\Longrightarrow x_1+x_2 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \mu$$

2-D Gaussian Distribution - Experimental Results

μ	1	1.5	2	2.5	3	3.5	4	4.5
Bayesian	75.985	85.485	91.93	96.13	98.335	99.34	99.81	99.95
CQS	75.985	85.485	91.93	96.13	98.335	99.34	99.81	99.95

Table: Classification of Normally Distributed 2-D Classes by the CQS 2-QS.

Multi-dimensional Gaussian Distribution

For the d-dimensional identical and symmetric Gaussian distributions, the classifier is:

$$x_1 + x_2 + \ldots + x_d = \frac{d}{2} \cdot \mu.$$

CQS/Dual CQS attains Bayes’ bound.

Rayleigh Distribution

$$f(X,\sigma) = \frac{x_1}{\sigma_{11}^2} e^{\frac{-x_1^2}{2\sigma_{11}^2}} \cdot \frac{x_2}{\sigma_{12}^2} e^{\frac{-x_2^2}{2\sigma_{12}^2}}, \quad -\infty < x_1 < \infty, -\infty < x_2 < \infty$$

$$f(X-\theta,\sigma) = \frac{x_1-\theta_1}{\sigma_{21}^2} e^{\frac{-(x_1-\theta_1)^2}{2\sigma_{21}^2}} \cdot \frac{x_2-\theta_2}{\sigma_{22}^2} e^{\frac{-(x_2-\theta_2)^2}{2\sigma_{22}^2}}, \quad -\infty < x_1 < \infty, -\infty < x_2 < \infty$$

CQS Positions:

$$q_1 = \left[\sigma_{11}\sqrt{2\,ln(3)},\ \sigma_{12}\sqrt{2\,ln(3)}\right]^T$$

$$q_2 = \left[\theta_1 + \sigma_{21}\sqrt{2ln\left(\frac{3}{2}\right)},\ \theta_2 + \sigma_{22}\sqrt{2ln\left(\frac{3}{2}\right)}\right]^T$$

2D Rayleigh Distribution-Pictorial Representation

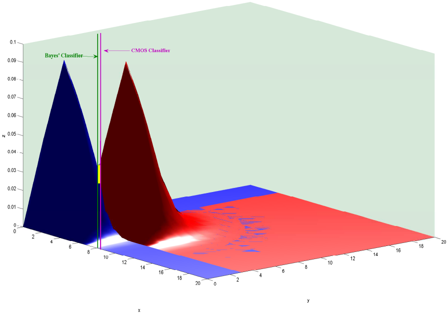


Figure: The upper bound of the differences of the error probabilities.

2-D Rayleigh Distribution - Experimental Results

No.	Order(n)	Moments	$\theta = 2$	$\theta = 1.5$	$\theta = 1.3$	$\theta = 1.2$
1	Median	$\left[\left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, \frac{1}{2}\right)\right]$	98.25	95.3	93.15	89.75
2	Two	$\left[\left(\frac{2}{3}, \frac{1}{3}\right), \left(\frac{2}{3}, \frac{1}{3}\right)\right]$	98.35	95.3	92.95	89.6
3	Four	$\left[\left(\frac{4}{5}, \frac{1}{5}\right), \left(\frac{4}{5}, \frac{1}{5}\right)\right]$	98.65	95.3	92.7	90.6
4	Four	$\left[\left(\frac{3}{5}, \frac{2}{5}\right), \left(\frac{3}{5}, \frac{2}{5}\right)\right]$	98.25	95.2	93.15	89.6
5	Six	$\left[\left(\frac{6}{7}, \frac{1}{7}\right), \left(\frac{6}{7}, \frac{1}{7}\right)\right]$	98.75	95.15	92.2 (D)	90.35 (D)
6	Six	$\left[\left(\frac{5}{7}, \frac{2}{7}\right), \left(\frac{5}{7}, \frac{2}{7}\right)\right]$	98.45	95.2	92.85	90.05
7	Six	$\left[\left(\frac{4}{7}, \frac{3}{7}\right), \left(\frac{4}{7}, \frac{3}{7}\right)\right]$	98.25	95.2	93.15	89.75
8	Eight	$\left[\left(\frac{8}{9}, \frac{1}{9}\right), \left(\frac{8}{9}, \frac{1}{9}\right)\right]$	98.8	95.2 (D)	92.25 (D)	89.9 (D)
9	Eight	$\left[\left(\frac{7}{9}, \frac{2}{9}\right), \left(\frac{7}{9}, \frac{2}{9}\right)\right]$	98.65	95.3	92.85	90.45
10	Eight	$\left[\left(\frac{5}{9}, \frac{4}{9}\right), \left(\frac{5}{9}, \frac{4}{9}\right)\right]$	98.25	95.25	93.1	89.7

Table: Classification of Rayleigh Distributed 2-D Classes by the CQS k-QS.

Ultimate PRS - Introduction

- Push the limit of PRS to its maximum extreme
- Aims to select/create a “single” prototype
- Tested on artificial and real-life data sets
- Attains comparable accuracy when compared to NN, NB and SVM
- Philosophy: *Use QS-based strategy to get PRS*

Ultimate PRSs: Histogram Based

- Vector-based Selective PRS - Histogram, Vector computations
  - Estimate Histogram in each dimension of vector space;
  - Get QS-based points in *all* dimensions;
  - Comparisons are vector-based.
- Scalar-based Selective PRS - Histogram, Scalar computations
  - Estimate Histogram in each dimension of vector space;
  - Get QS-based points in *all* dimensions;
  - Scalar Comparisons, Majority vote.

Ultimate PRSs: Naive Bayes Approach

- Vector-based Creative PRS - Naive Bayes, Vector computations
  - Assuming Gaussian, estimate CQS positions in each dimension;
  - Create a new vector pattern with the scalar components;
  - Comparisons are vector-based.
- Scalar-based Creative PRS - Naive Bayes, Scalar computations
  - Assuming Gaussian, estimate CQS positions in each dimension;
  - Perform scalar comparisons in each dimension;
  - Majority vote to determine the identity.

Ultimate PRSs: Single Feature of a Single Pattern

- Ultimate PRS: Single Feature of a Single Pattern
  - Identify the best feature in the training phase;
  - Assuming Gaussian, estimate the CQS positions of the selected feature;
  - Perform scalar comparison to determine the identity.

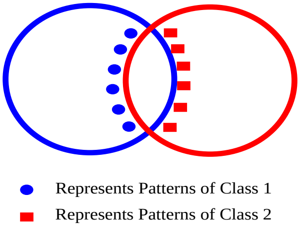
Experimental Results

Data set	Traditional Classifiers			CQS Classifier				
	NB	NN	SVM	Vector		Scalar		SFSP
				NB	Histogram	NB	Histogram	
WOBC	96.40	96.60	96.99	96.94	95.06	94.35	92.06	93.04
WDBC	92.97	96.66	97.71	93.43	90.07	89.25	86.82	91.29
Diabetes	73.11	71.90	73.84	73.76	65.74	76.74	43.41	73.63
Hepatitis	83.19	82.58	84.54	76.67	75.13	81.87	81.00	83.93
Iris	95.13	96.00	96.67	94.4	92.50	93.80	77.80	95.50
AU Credit	87.40	85.90	85.51	94.76	84.21	83.03	48.19	84.84
Heart	83.00	84.40	85.60	84.59	83.93	77.11	60.67	78.52
Vote	94.29	90.23	94.33	93.43	91.0	89.10	85.36	95.40

Table: Classification of Real-life data sets by CQS.

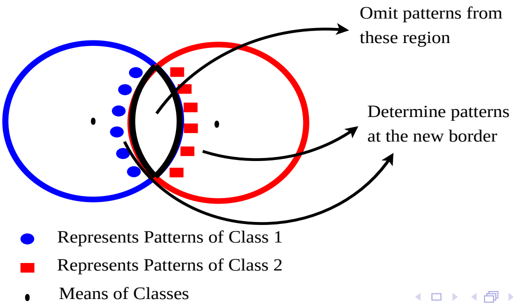
"Anti-Bayesian" BI

- Intent: Utilize QS-based philosophy
- Propose "Anti-Bayesian" BI
- Suggests a new definition for "border pattern"
- Border patterns: Close to *neither* the mean nor the class boundary



ABBI Algorithm

- Omit patterns that fall at the overlapped region
- Determine the patterns at the new border
- These are the “PRS” patterns
- Perform classification based on a selected number of new border patterns



Separated classes

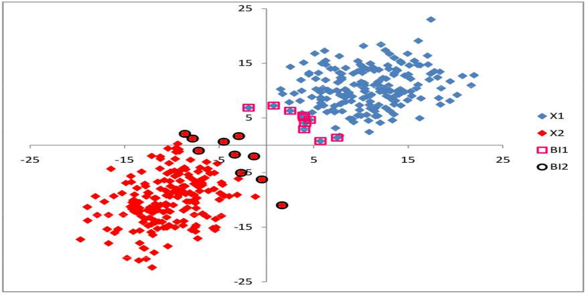


Figure: Border patterns: Fairly separable classes obtained by ABBI.

Semi-overlapped classes

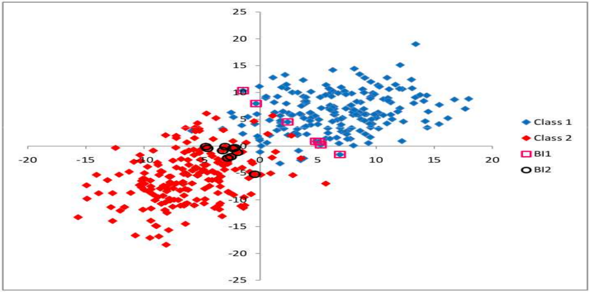


Figure: Border patterns: Semi-overlapped classes obtained by ABBI.

Overlapped classes

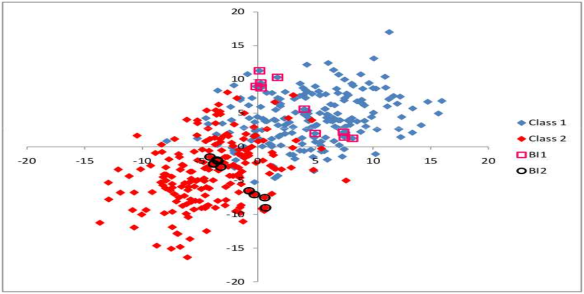


Figure: Border patterns:Overlapped classes obtained by ABBI.

Experimental Results

Data set	kNN	NB	SVM	ABBI
WOBC	96.60	96.40	95.99	95.80
WDBC	96.66	92.97	97.71	92.39
Diabetes	75.26	73.1098	76.70	72.30
Hepatitis	82.58	83.19	82.51	80.27
Iris	95.13	96.00	96.67	94.53
Statlog (Australian Credit)	85.90	87.40	85.51	78.85
Statlog (Heart)	84.40	83.00	85.60	82.50
Vote	94.2857	90.23	94.33	90.76

Table: Classification of Real Data.

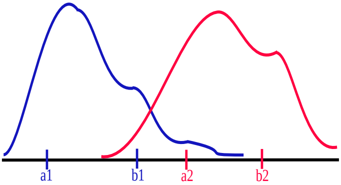
“Anti-Bayesian” Clustering

- Intent: Utilize QS-based philosophy for Clustering
- Propose “Anti-Bayesian” Clustering Schemes
- Use QS instead of the k-means concepts
- Top-Down and Bottom-Up Schemes Possible

Outline: “Anti-Bayesian” Clustering

- “Anti-Bayesian” Paradigm
- k-means Clustering Algorithm
- Clustering Algorithm: “Anti-Bayesian” Paradigm
- Experimental Results

Alternate View: Generic “Anti-Bayesian” Classifier

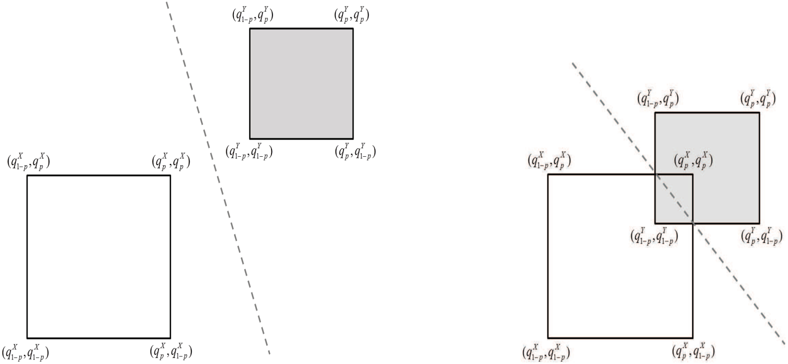


- a_1, b_1: QS for ω_1
- a_2, b_2: QS for ω_2

When test sample x^* comes:

- If $\mathbf{x}^* < \mathbf{b_1}$, $x^* \in \omega_1$
- If $\mathbf{x}^* > \mathbf{a_2}$, $x^* \in \omega_2$
- If $\mathbf{b_1} < \mathbf{x}^* < \mathbf{a_2}$, decision is based on $\|x^* - \mathbf{b_1}\|$ and $\|x^* - \mathbf{a_2}\|$
- The classification border is: $B = \frac{b_1 + a_2}{2} \implies \frac{q_p^{\omega_1} + q_{1-p}^{\omega_2}}{2}$

Generic Two-Class "Anti-Bayesian" Classifier



k-means Clustering Algorithm

- Suppose we want to group the following into two clusters

k-means Clustering Algorithm

- Choose two arbitrary starting positions (centroids)

k-means Clustering Algorithm

- Classify each point to the closest centroid

k-means Clustering Algorithm

- Update the centroid as the average in the cluster

k-means Clustering Algorithm

- Again classify each point to the closest centroid

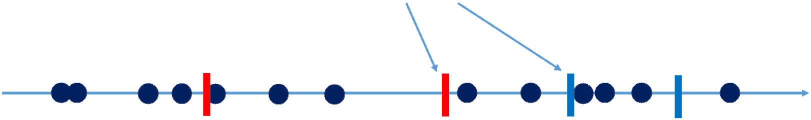
Repeat until Convergence

Clustering based on “Anti-Bayesian” Paradigm

- Here: Select arbitrary quantiles for the two clusters

Clustering based on “Anti-Bayesian” Paradigm

- Assign points to clusters by the average of these



Clustering based on “Anti-Bayesian” Paradigm

- Which give the result

Clustering based on “Anti-Bayesian” Paradigm

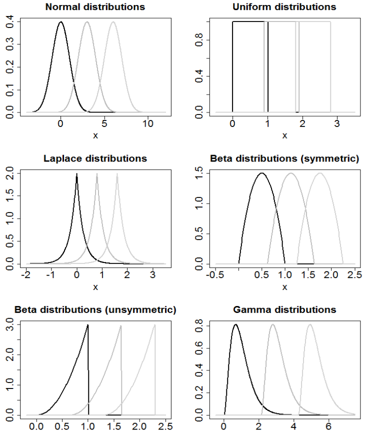
- Update quantiles by quantile estimator

Clustering based on “Anti-Bayesian” Paradigm

- Again assign points to clusters based on these quantiles

Repeat until Convergence

Experimental Results: 1D



- Create synthetic data with $N = 10$ and $N = 1000$ points in each cluster
- Three clusters
- 1D and 2D
- Since we know the cluster of each sample point, the number of errors can be measured
- Use the cluster labeling that minimizes the number of errors
- Repeat experiment many times to remove Monte Carlo error

Experimental Results: 1D

Portion of points in a wrong cluster

	$N = 10$		$N = 1000$	
Distribution	k-means	“Anti-Bayes”	k-means	“Anti-Bayes”
Normal	0.105	0.105	0.090	0.089
Uniform	0.116	0.104	0.106	0.157
Laplace	0.149	0.163	0.136	0.135
Beta (symmetric)	0.145	0.138	0.125	0.139
Beta (asymmetric)	0.081	0.087	0.074	0.098
Gamma	0.143	0.170	0.113	0.132

Width confidence intervals ≈ 0.002

Experimental Results: 2D

	$N = 10$		$N = 1000$	
Distribution	k-means	“Anti-Bayes”	k-means	“Anti-Bayes”
Normal	0.140	0.145	0.106	0.107
Uniform	0.102	0.102	0.074	0.078
Laplace	0.139	0.156	0.108	0.108
Beta (symmetric)	0.145	0.147	0.108	0.109
Beta (unsymmetric)	0.114	0.168	0.081	0.121
Gamma	0.141	0.202	0.102	0.124

Width confidence intervals ≈ 0.002

Experimental Results

- We can use the “Anti-Bayesian” classification framework to cluster data points
- The clustering algorithm performs remarkably accurate
- Future work
  - Higher dimensions
  - Real data
  - Generalize other clustering (Top-Down/Bottom-Up) algorithms that depend on the Bayesian philosophy

“Anti-Bayesian” Text Classification

- Intent: Utilize QS-based philosophy for Text Classification
- Propose “Anti-Bayesian” Text Classification Schemes

Outline: “Anti-Bayesian” Text Classification

- Complexity: Text Classification
- Traditional Methods
- Text Classification: “Anti-Bayesian” Paradigm
- Text Classification: Experimental Results

Complexity: Text Classification

- Classes: Text Classification
  - Different Classes of Textual Data
  - May be presented as Files, News Items, Test Messages, Word-of-Mouth
- Features: Text Classification
  - Syntactic and not Statistical
  - Statistical features extracted from these
  - Classification achieved using statistical features
  - Bayes, Naïve-Bayes, SVM etc.: Using these statistical features

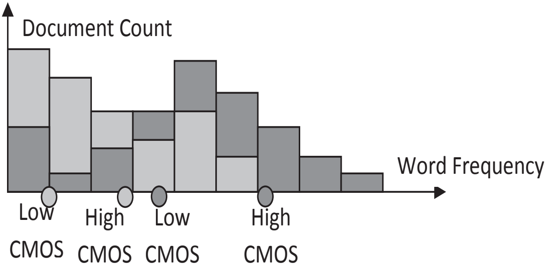
Complexity: Text Classification

- Representation: Text Classification
  - Bag of Words (for example)
  - Remove “Stop Words”
  - Use Term Frequencies (TF) in the documents
    - Importance of term; WRT the document
  - Use Term Frequency-Inverse Document Frequencies (TFIDF)
    - Importance term: WRT the document and WRT the Corpus
    - Weights a term: How many times it is in a document
    - Weight negatively biased: No. of documents term found in
  - Use Syntactic (not Statistical) similarities (Cosine Similarity metric)

Text Classification: "Anti-Bayesian" Paradigm

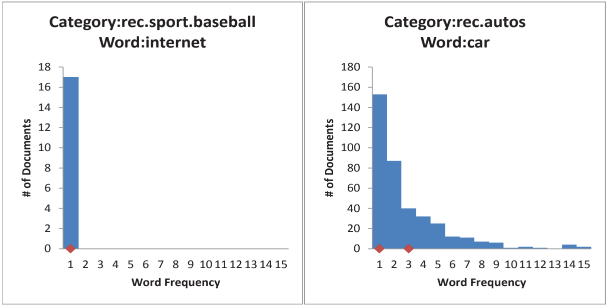
- What is traditionally done?
  - Traditionally: Classify Sports Vs. Business
  - Use terms typical for both: Basketball Vs. Dollar
- What do we want to do?
  - Traditionally: Classify Sports Vs. Business
  - Do something Bizarre
  - Use terms that are atypical for both classes
  - Use Distant Quantiles of the TF distributions

Text Classification: QS-based Features



QS-based features: Histograms of lower (light) & higher classes (dark); QS points

Text Classification: QS-based Features



Histograms & $\frac{1}{3}$ and $\frac{2}{3}$ QS points for “internet ” and “car” from “baseball” & “autos”

Data Files and Metrics Used

The Data Files Used:

comp.graphics	alt.atheism	sci.crypt	misc.forsale
comp.sys.mac.hardware	talk.religion.misc	sci.electronics	rec.autos
comp.windows.x	talk.politics.guns	sci.med	rec.motorcycles
comp.os.ms-windows.misc	talk.politics.mideast	sci.space	rec.sport.hockey
comp.sys.ibm.pc.hardware	talk.politics.misc	soc.religion.christian	rec.sport.baseball

The “20-Newsgroups” used in the experiments.

The Metric Used:

$$P_i = \frac{TP_i}{TP_i + FP_i}; \qquad R_i = \frac{TP_i}{TP_i + FN_i}$$

$$F_i = \frac{2P_iR_i}{P_i + R_i}; \qquad macro - F1 == \frac{1}{20}\sum_{i=1}^{20} F_i$$

Results: 1

Classifier	CMQS Points	*macro-F1* Score
“Anti-Bayesian”	1/2, 1/2	0.709
	1/3, 2/3	0.662
	1/4, 3/4	0.561
	1/5, 4/5	0.465
	2/5, 3/5	0.700
	2/7, 5/7	0.611
	3/7, 4/7	0.710
	3/8, 5/8	0.686
	2/9, 7/9	0.515
	4/9, 5/9	0.713
	3/10, 7/10	0.631
BOW		0.604
BOW-TFIDF		0.769
Naïve-Bayes		0.780

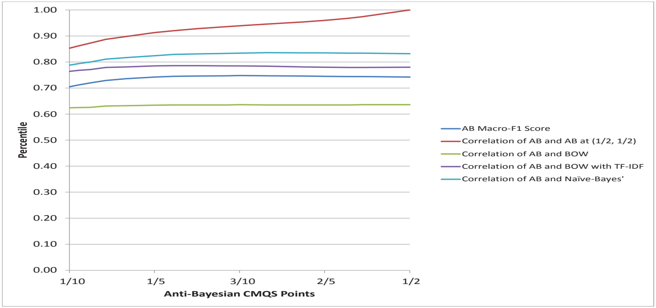
Macro-F1 score for 100 classifications; “Anti-Bayesian” used the TF features

Results: 2

Classifier	CMQS Points	*macro-F1* Score
“Anti-Bayesian”	1/2, 1/2	0.742
	1/4, 3/4	0.746
	2/5, 3/5	0.745
	1/6, 5/6	0.736
	1/7, 6/7	0.729
	2/7, 5/7	0.747
	3/7, 4/7	0.744
	3/8, 5/8	0.746
	2/9, 7/9	0.745
	4/9, 5/9	0.744
	1/10, 9/10	0.705
	3/10, 7/10	0.748
BOW		0.604
BOW-TFIDF		0.769
Naïve-Bayes		0.780

Macro-F1 score for 100 classifications; “Anti-Bayesian” used the TFIDF features

Correlation of Classifiers



Correlation between the classifiers, where the “Anti-Bayesian” used the TFIDF features

Summary of Results

- For *all* CMQS pairs, the “Anti-Bayesian” much better than BOW
- For example: BOW score of 0.604, CQMS score for $\langle \frac{1}{3}, \frac{2}{3} \rangle$, was 0.747
- For $\langle \frac{1}{4}, \frac{3}{4} \rangle$, $\langle \frac{3}{7}, \frac{4}{7} \rangle$ and $\langle \frac{4}{9}, \frac{5}{9} \rangle$: consistently higher – 0.746, 0.744 and 0.744
- Classifiers are quite “uncorrelated”; Fusion techniques possible!!!
- Counter-intuitive: Characterize documents by non-central syntactic features

Conclusions

- Provided foundational theory for QS-based PR
- Proposed a strategy named CQS for parametric PR
- CQS attained optimal Bayes’ bound for symmetric distributions
- CQS attained near-optimal Bayes’ bound for asymmetric distributions
- Proposed Ultimate PRS – one which had only *a single* sample per class
- Provided rationale for BI algorithms
- Proposed a BI algorithm based on the new “Anti-Bayesian” concept of borders
- Introduced phenomenon of “Anti-Bayesian” Clustering
- Pioneered “Anti-Bayesian” Text Classificatiom

Future Work

- Multi-class Problems
- Non-parametric Distributions
- Kernel-based Methods
- Decision Trees

Actually: Whole avenues of PR are now open for research!

Thank You!

Open for Questions...