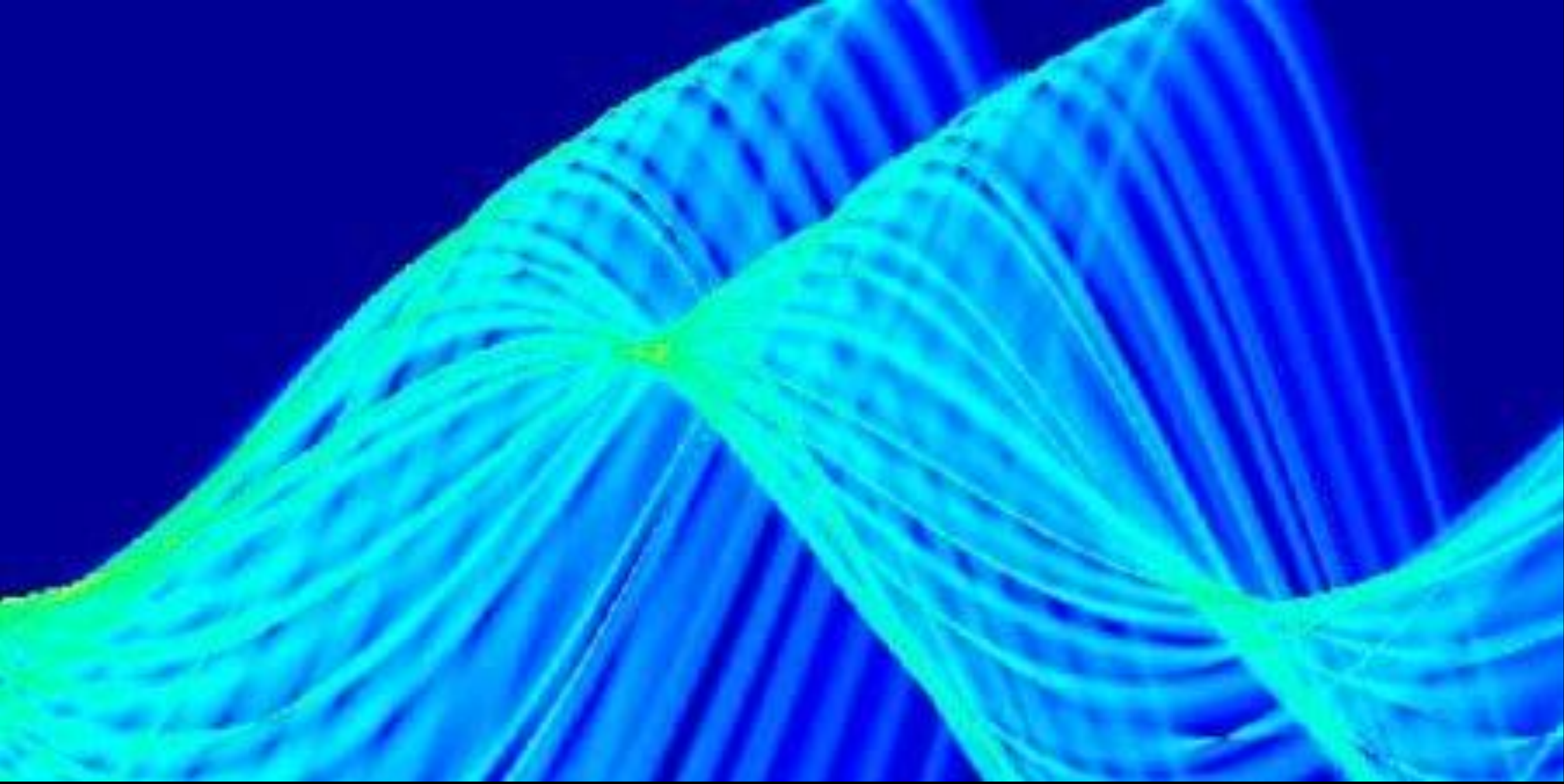


A Really Brief Introduction to Pattern Classification for Acousticians

Colin Jemmott

~~February 2008~~
December 2019

Hypothesis Testing

Who is speaking?

Was there an earthquake? Where?

Is that a bad sensor?

Is that a car or a tank driving down the road?

Is that a good weld?

When will this gearbox fail?

Is the recording a marine mammal or a cargo ship?

Pattern Classification Overview

Automation of these decisions for cost or repeatability

Also known as Decision Theory, Hypothesis Testing, Machine Learning, etc.

Speaker Recognition

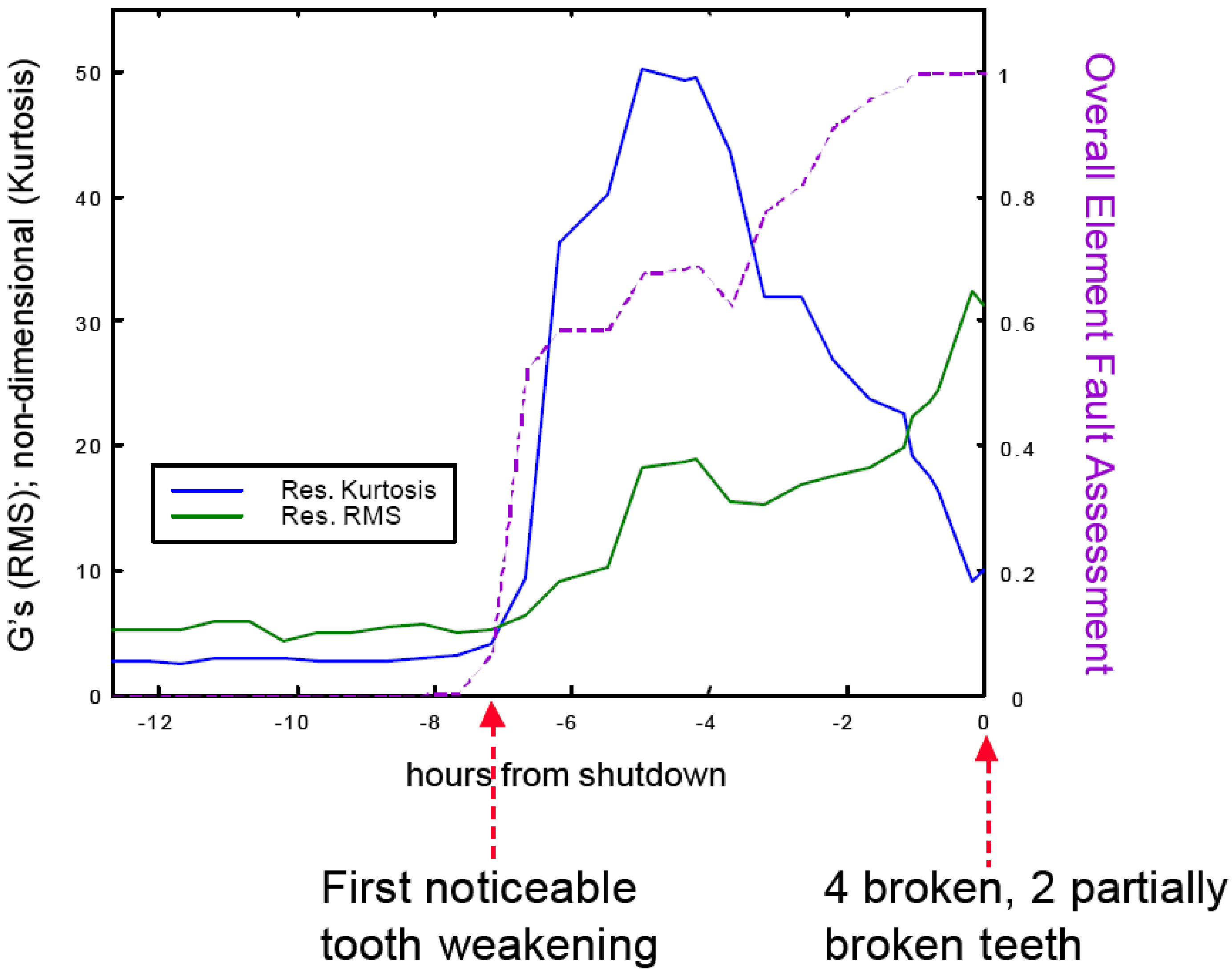
Receive Data – digitized acoustic time series (audio file)

Extract Features – Formant frequencies, speaking rate, etc.

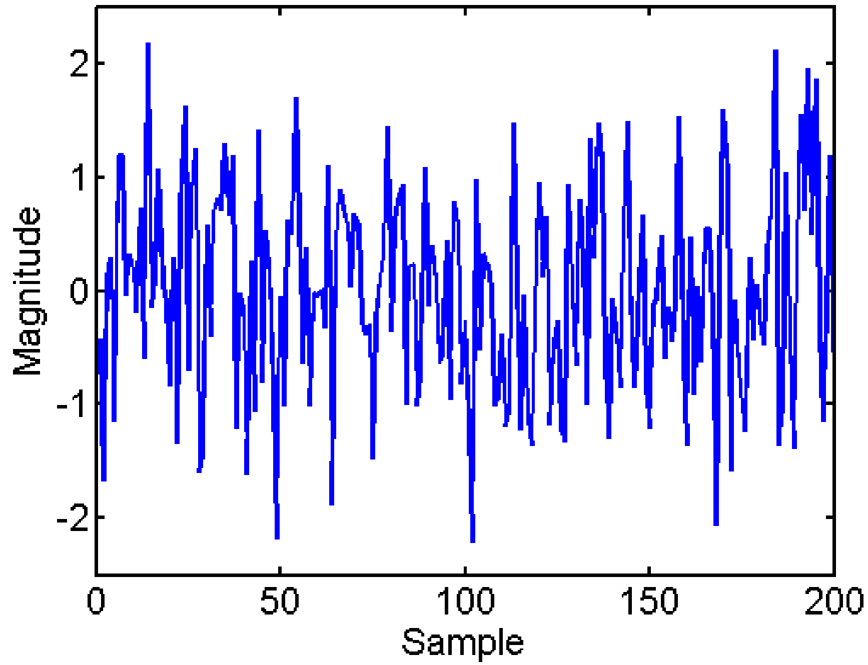
Assign to a class – Based on the values of the features, pick the most likely candidate speaker

"Application of sensor fusion and signal classification techniques in a distributed condition monitoring system"

Karl Reichard, Mike Van Dyke, Ken Maynard



Is the recording a marine mammal or a cargo ship?

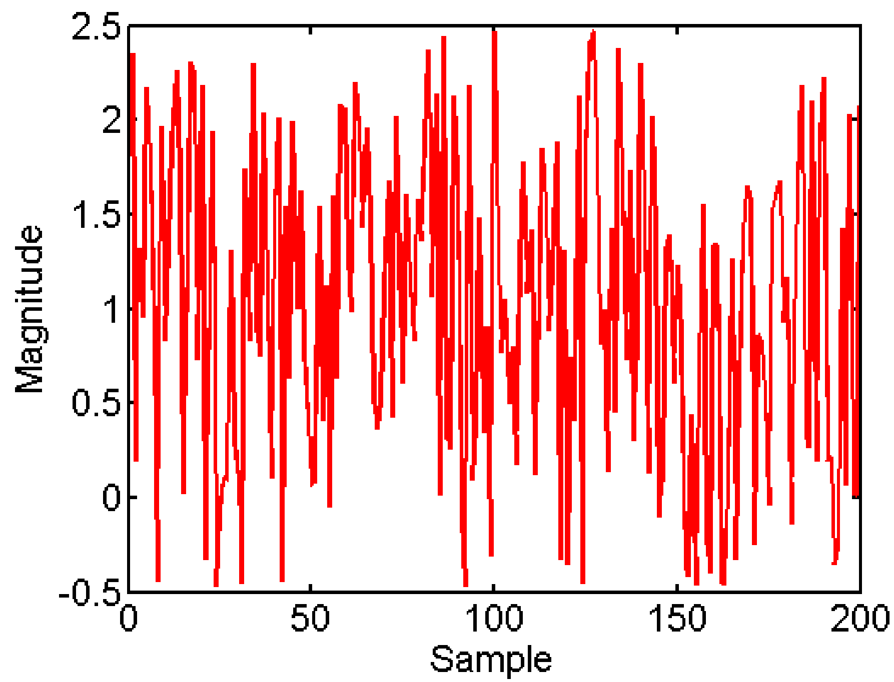


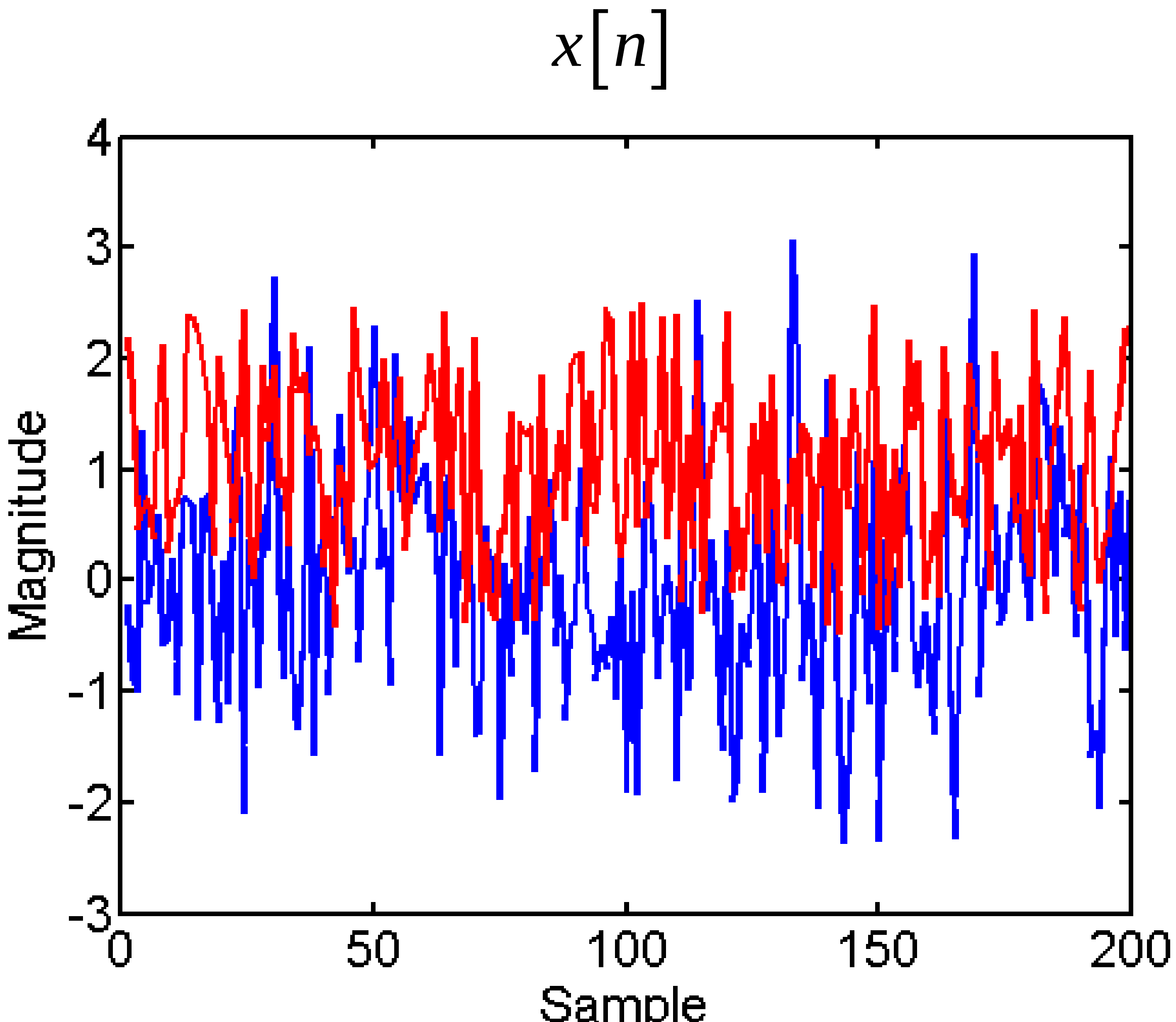
Hypothesis 1:

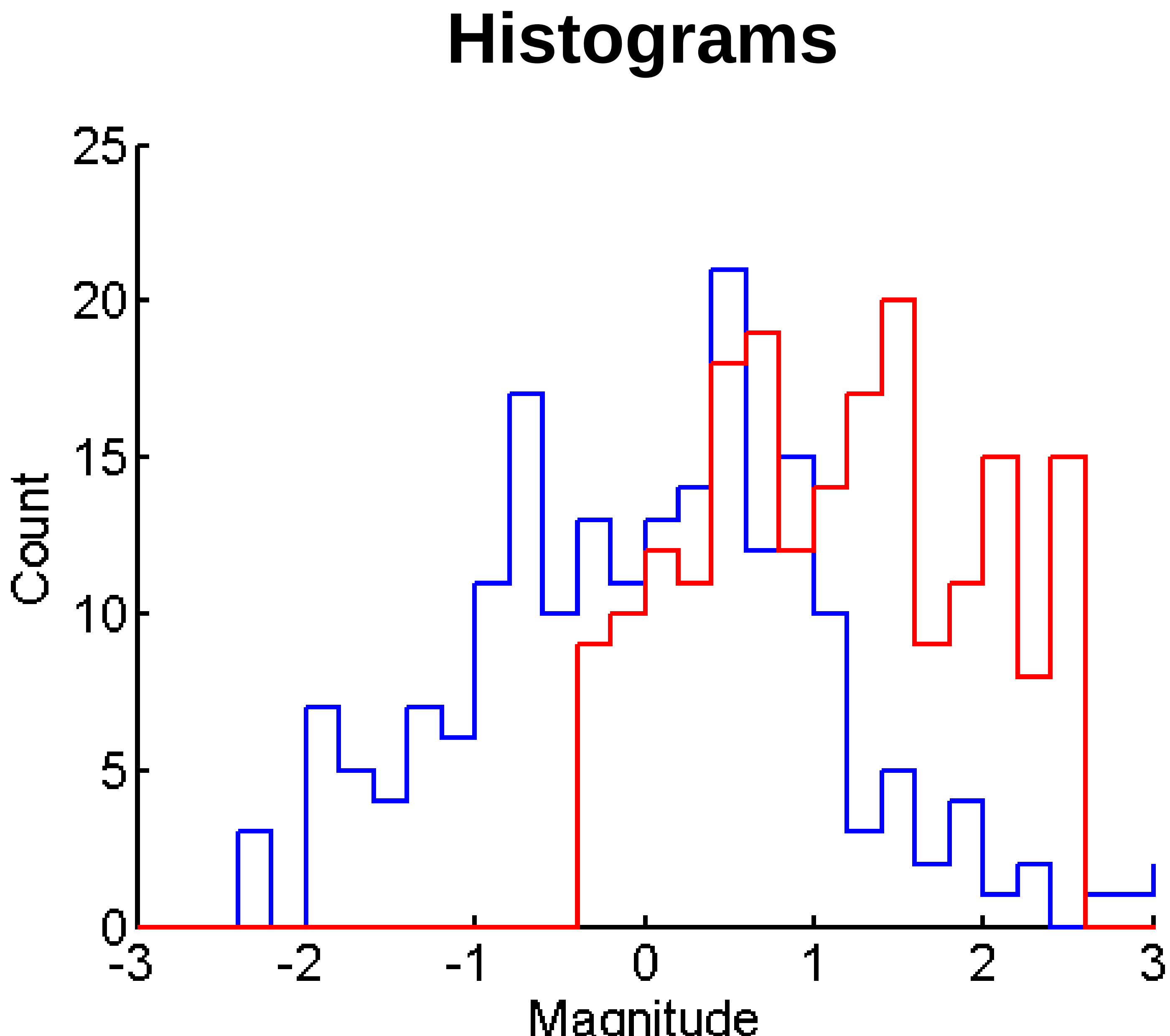
Marine Mammal

Hypothesis 2:

Cargo Ship



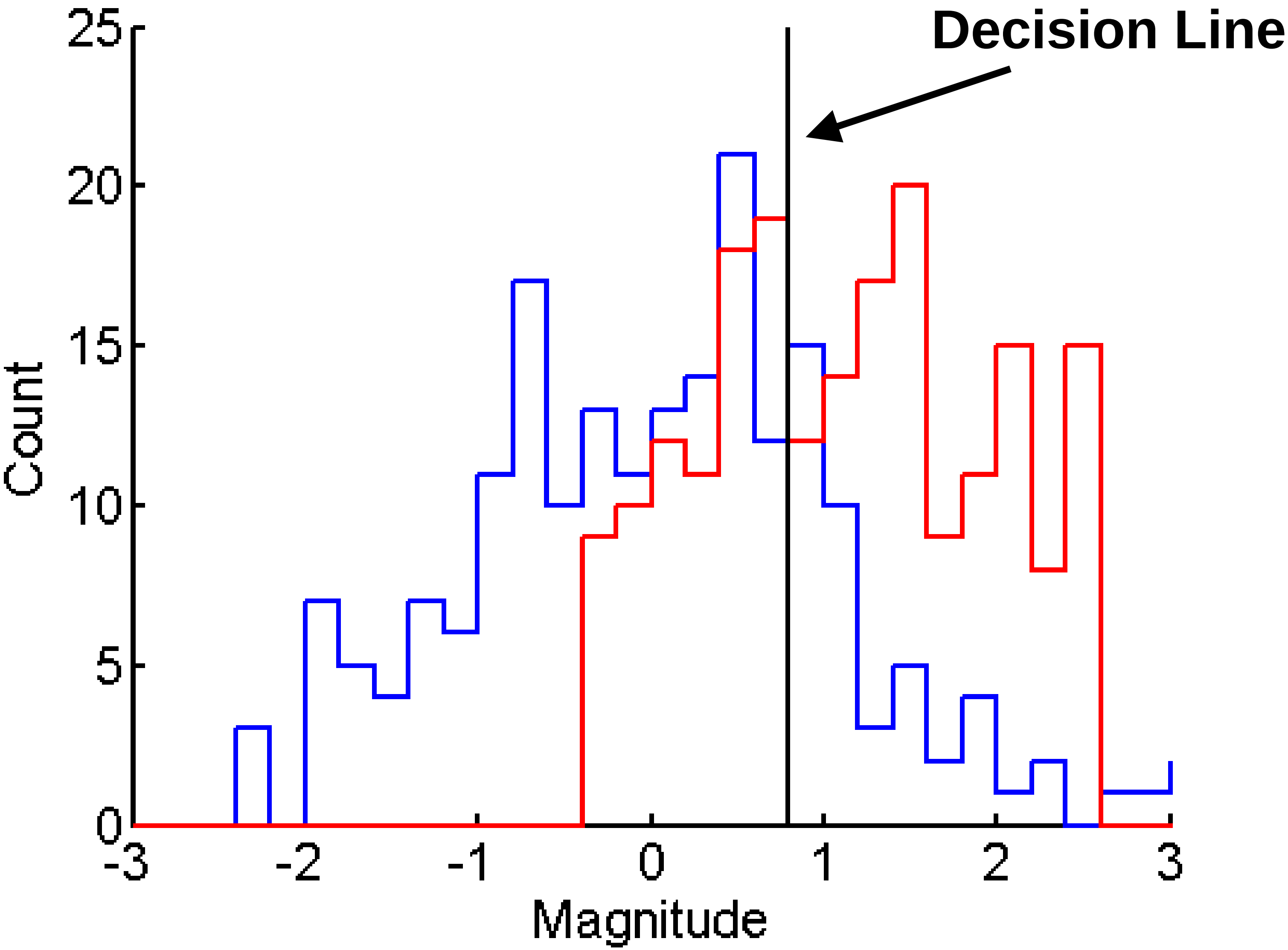

$x[n]$
Magnitude
Sample
4
3
2
1
0
-1
-2
-3
0
50
100
150
200



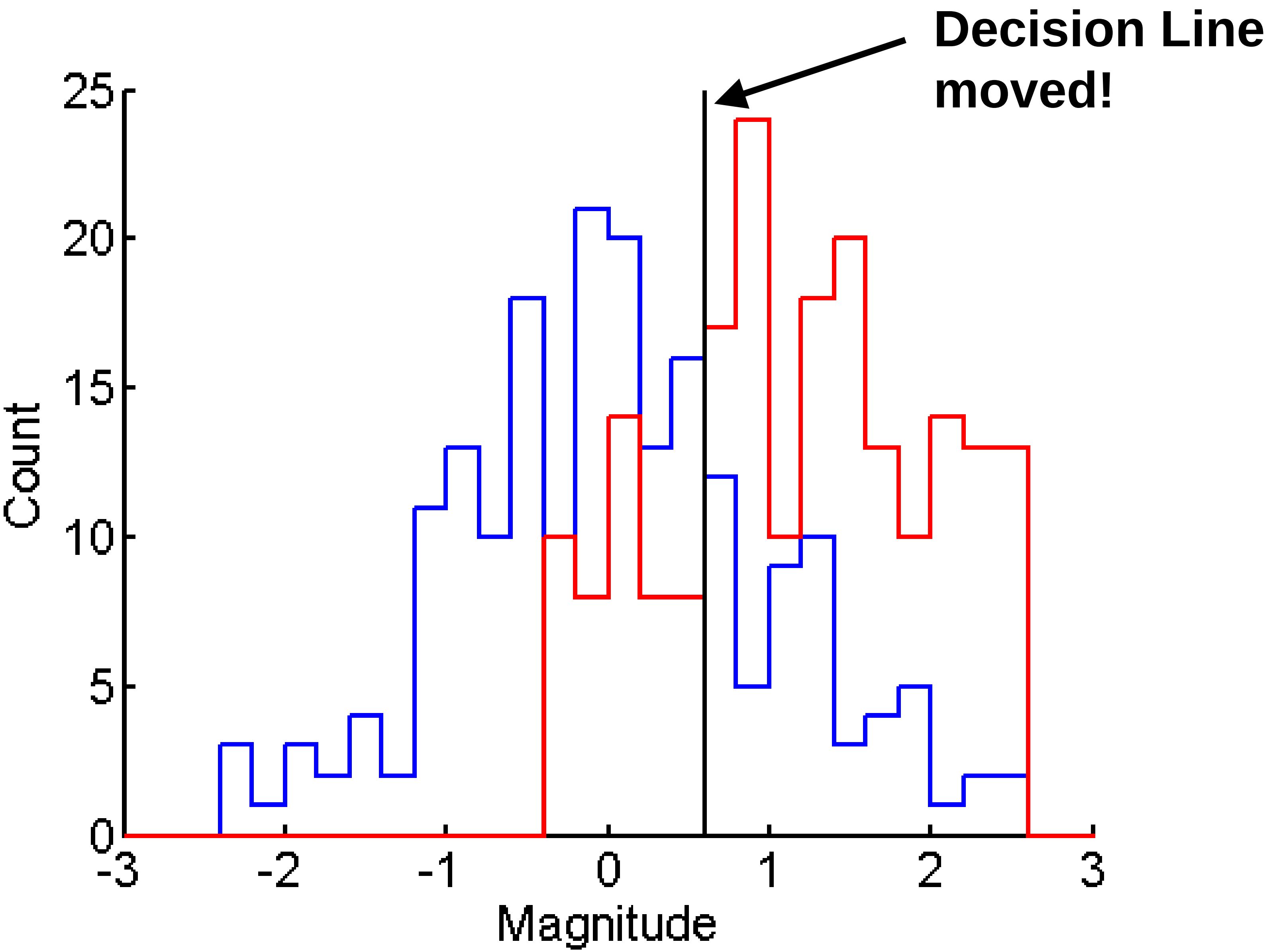
Histograms
Count
Magnitude
0
5
10
15
20
25
-3
-2
-1
0
1
2
3

$$\text{If} \quad x[n] \le 0.8 \quad \text{then guess} \quad H_1 \quad \text{marine mammal}$$

$$\text{If} \quad x[n] > 0.8 \quad \text{then guess} \quad H_2 \quad \text{cargo ship}$$



Histograms are very sensitive to bin location and width



There *is* a rigorous way to test hypotheses

$p\left(H_1 \mid x\right)$ **Probability that H_1 is true, given that x has happened**

$p\left(H_2 \mid x\right)$ **Probability that H_2 is true, given that x has happened**

We want to know which hypothesis is more likely

If $p\left(H_1 \mid x\right) > p\left(H_2 \mid x\right)$ **guess H_1 is true**

If $p\left(H_1 \mid x\right) < p\left(H_2 \mid x\right)$ **guess H_2 is true**

We write this as $p(H_1 \mid x) \underset{<}{\overset{>}{}} p(H_2 \mid x)$

If we assume both hypotheses are equally likely,

$$p(H_1) = p(H_2) = \tfrac{1}{2}$$

We can find

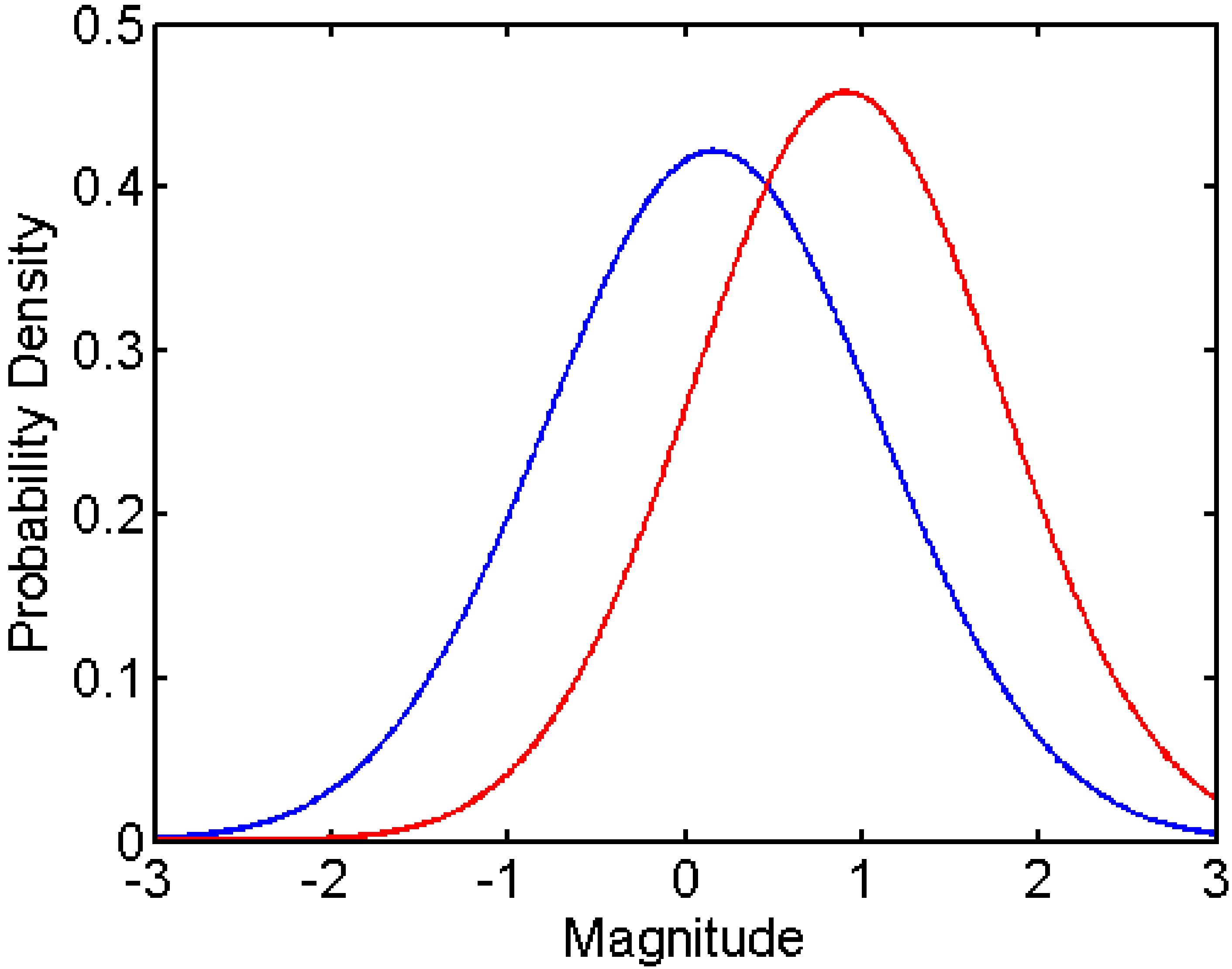
$$\boxed{p(x \mid H_1) \underset{<}{\overset{>}{}} p(x \mid H_2)}$$

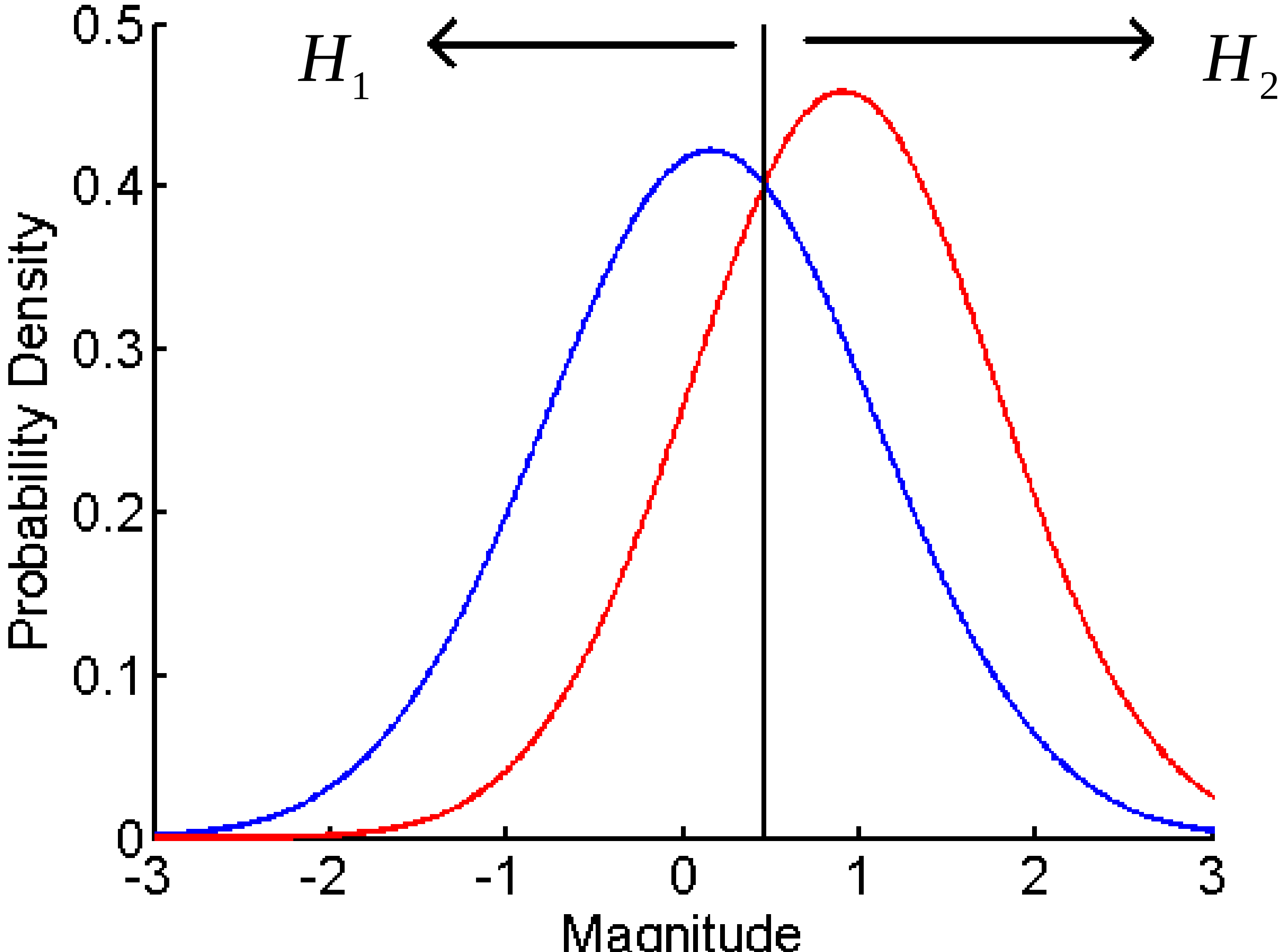
$p(x \mid H_1)$ Probability of event x happening, given that H_1 is true

$p(x \mid H_2)$ Probability of event x happening, given that H_2 is true

Assume $p\left(x \mid H_{1,2}\right)$ follows a Gaussian Distribution

(aka: normal distribution, bell curve)

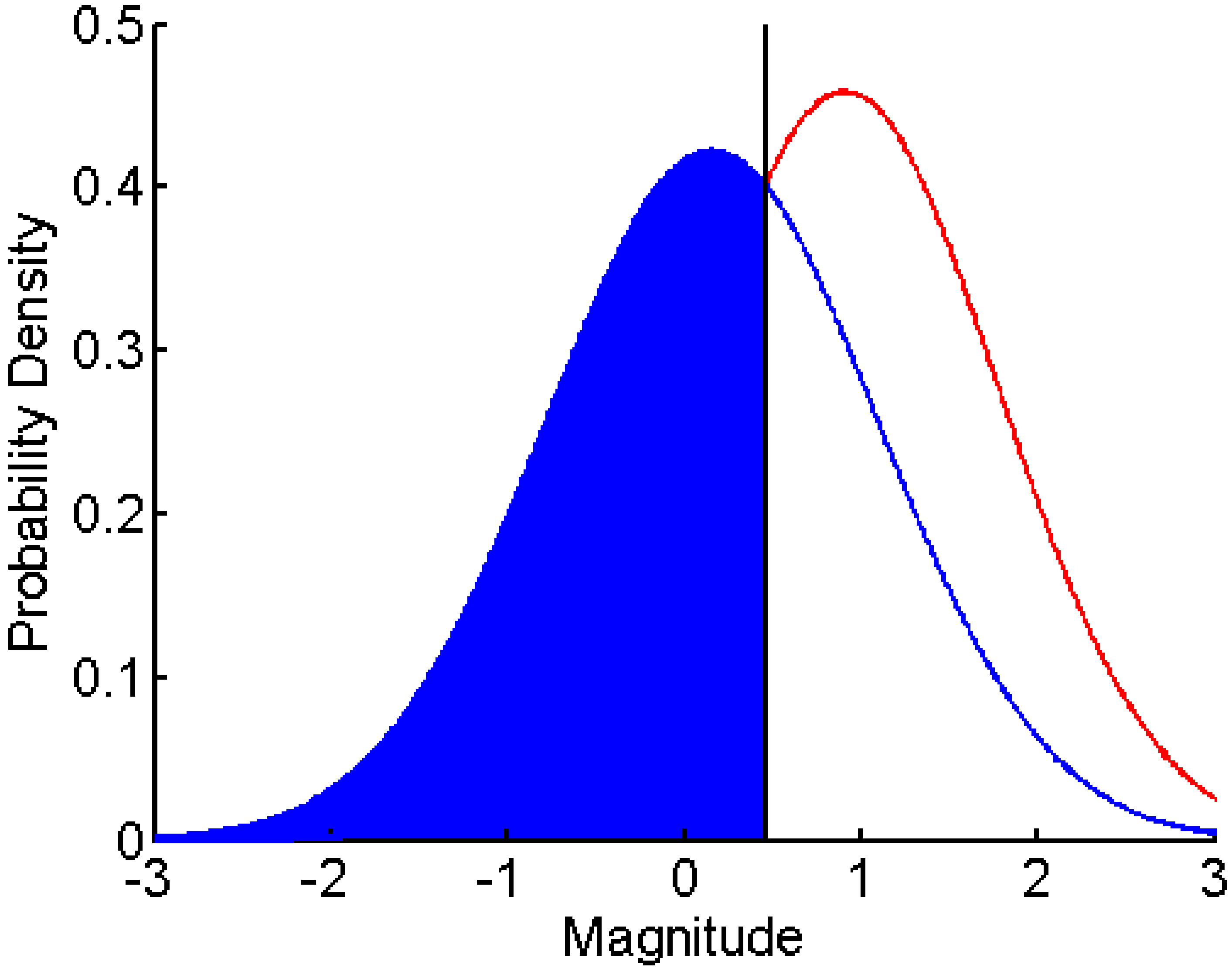




H_1
H_2
0.5
0.4
0.3
0.2
0.1
0
Probability Density
-3
-2
-1
0
1
2
3
Magnitude

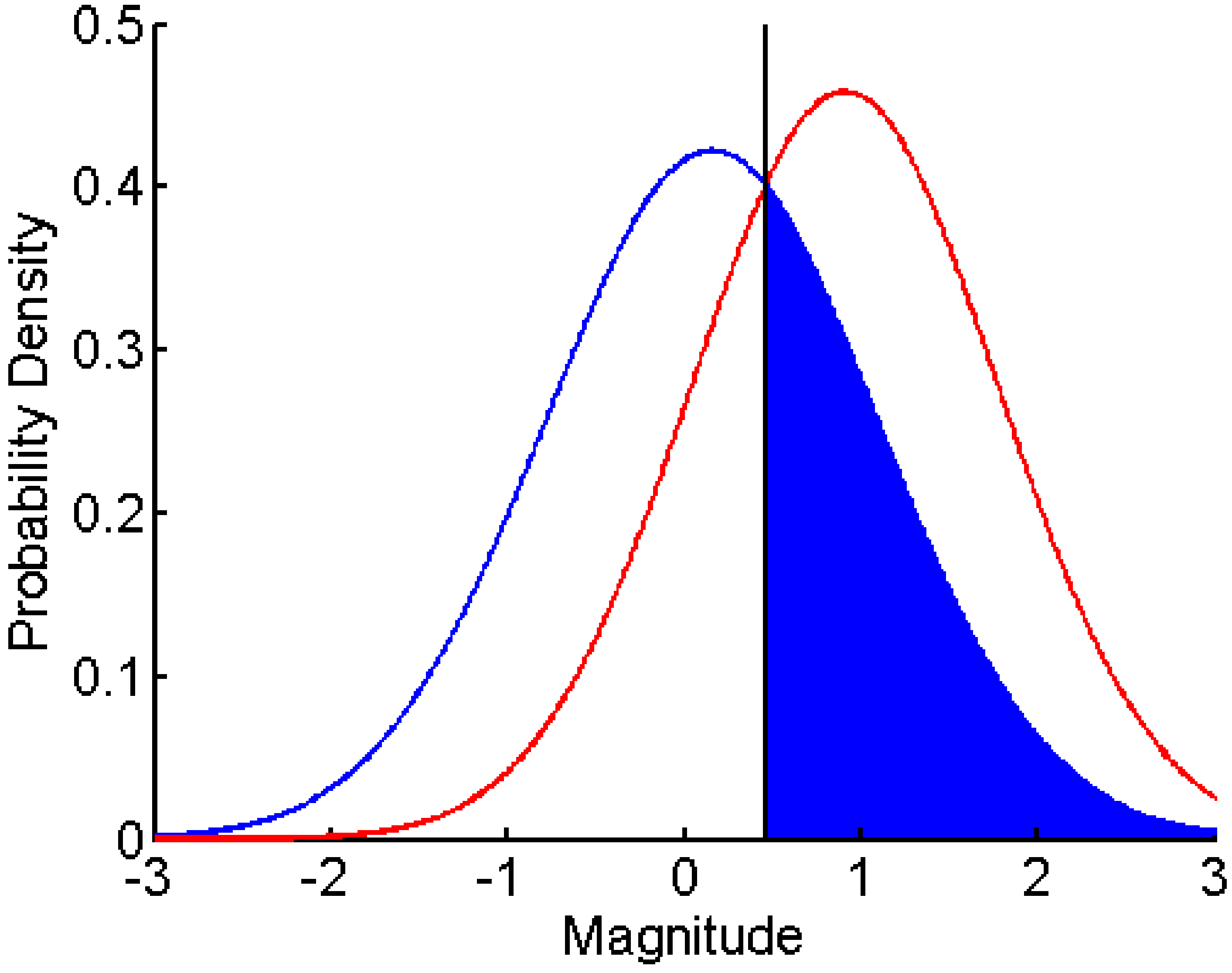
Correct Marine Mammal Classification

Cost is C_{11}



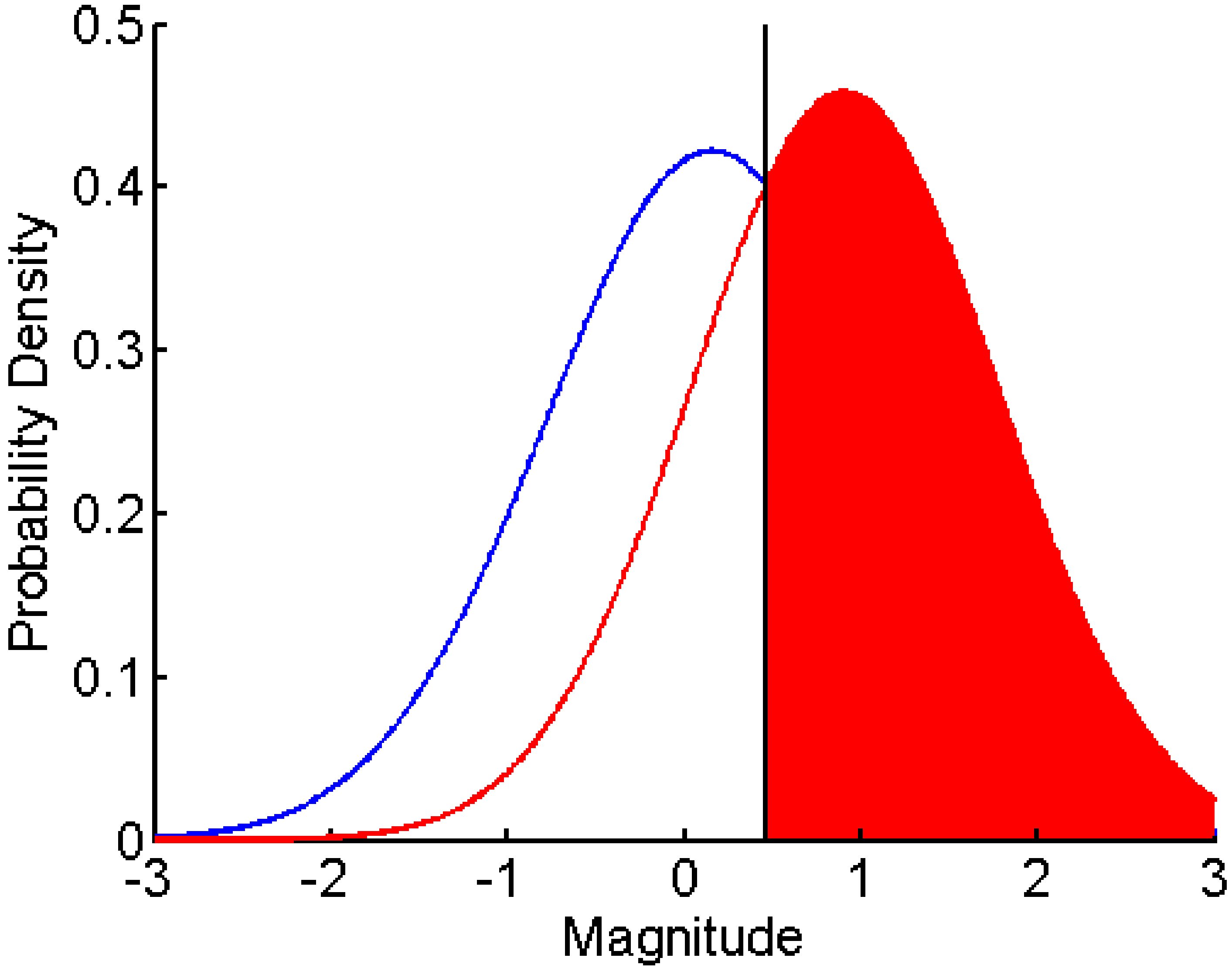
Error, you called a marine mammal a cargo ship!

Cost is C_{12}



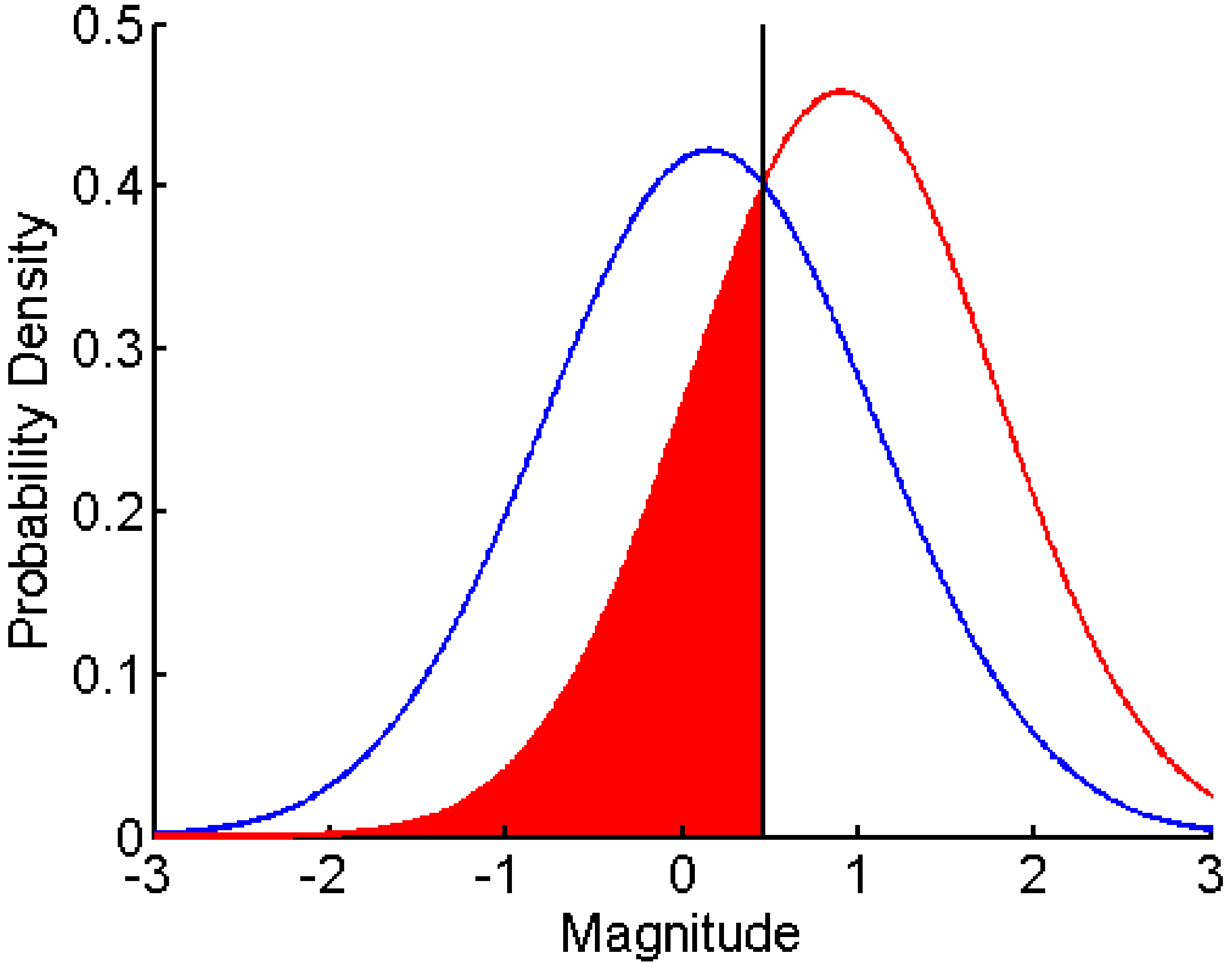
Correct Cargo Ship Classification

Cost is C_{22}



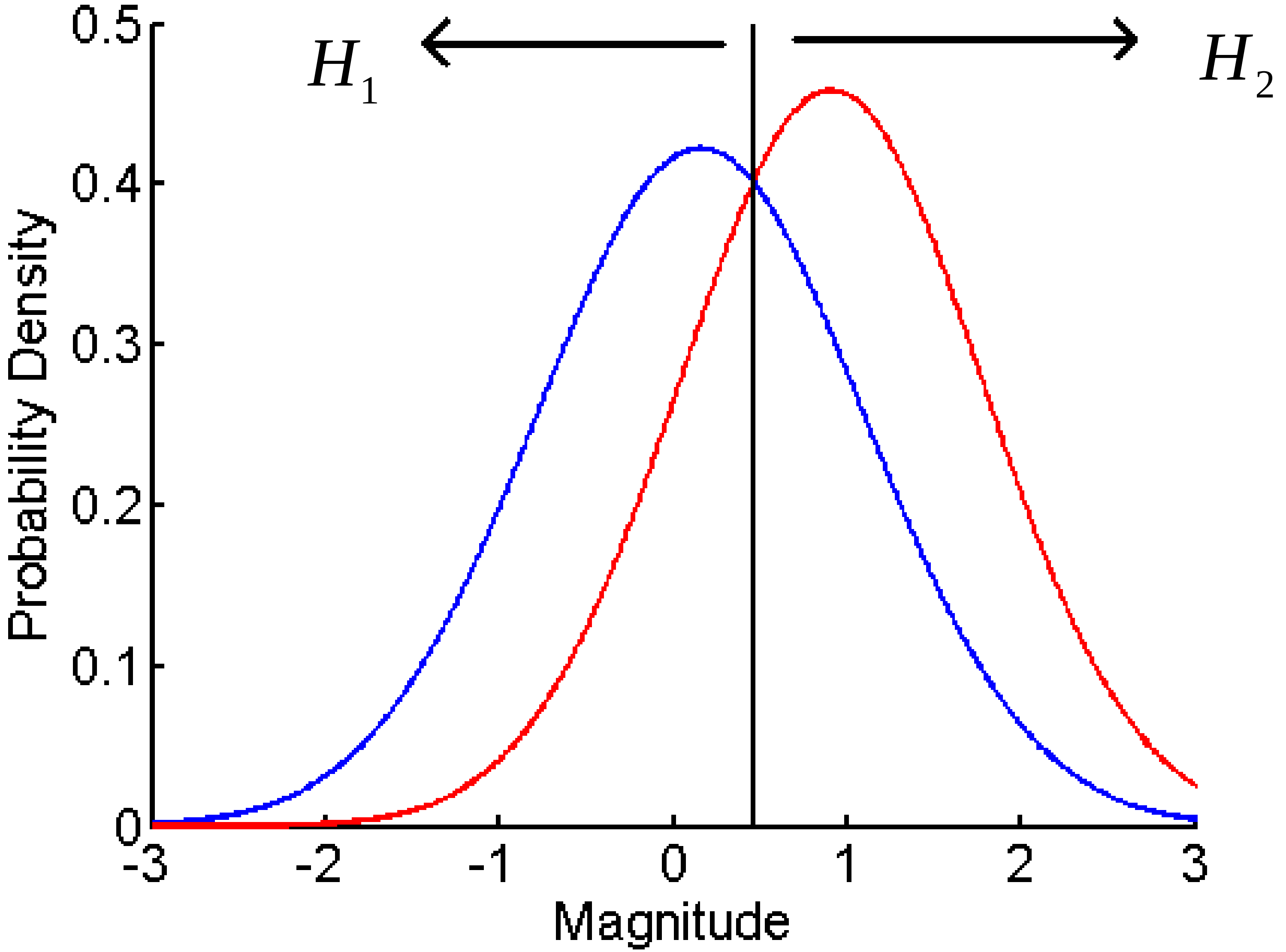
Error, you called a cargo ship a marine mammal!

Cost is C_{21}

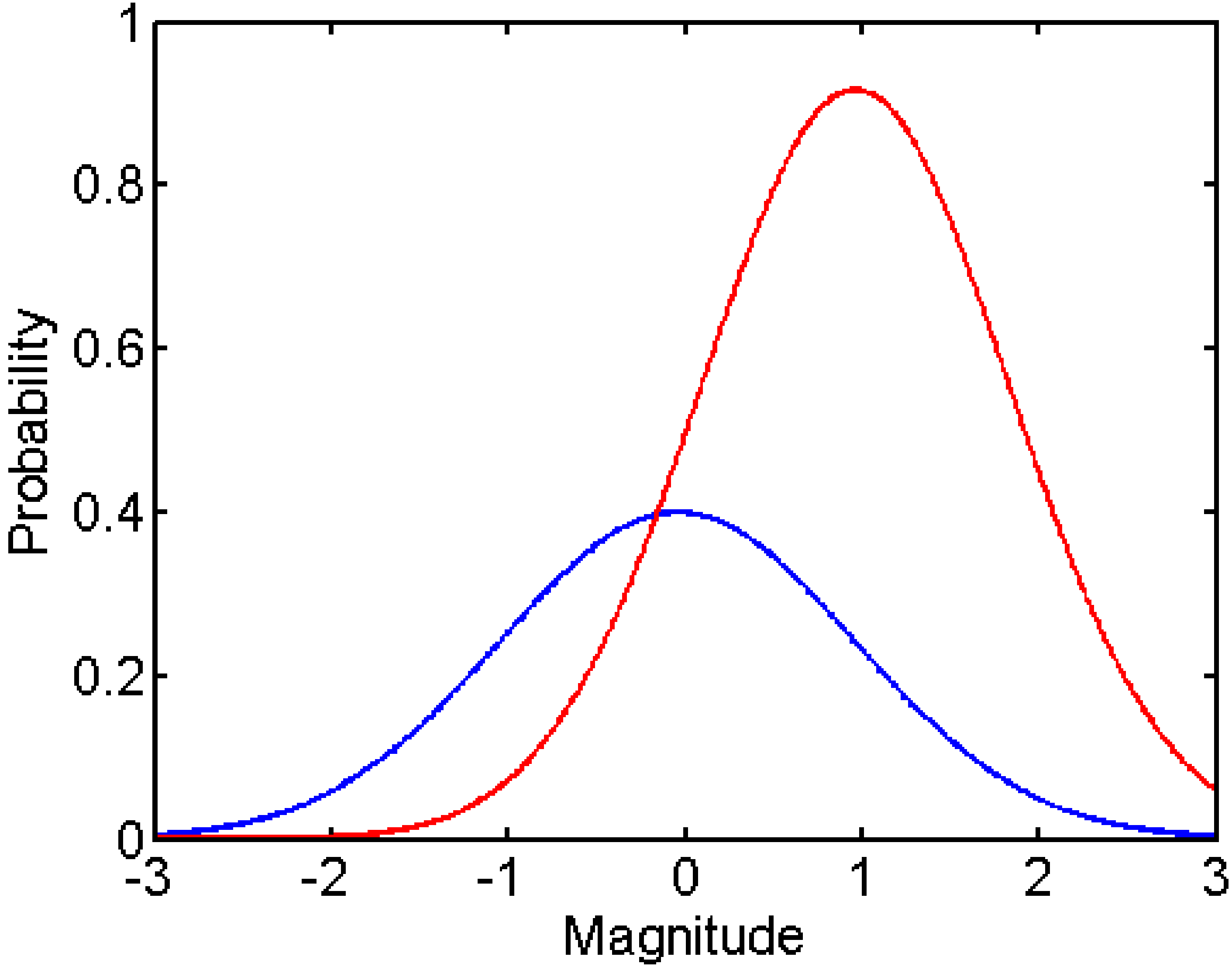


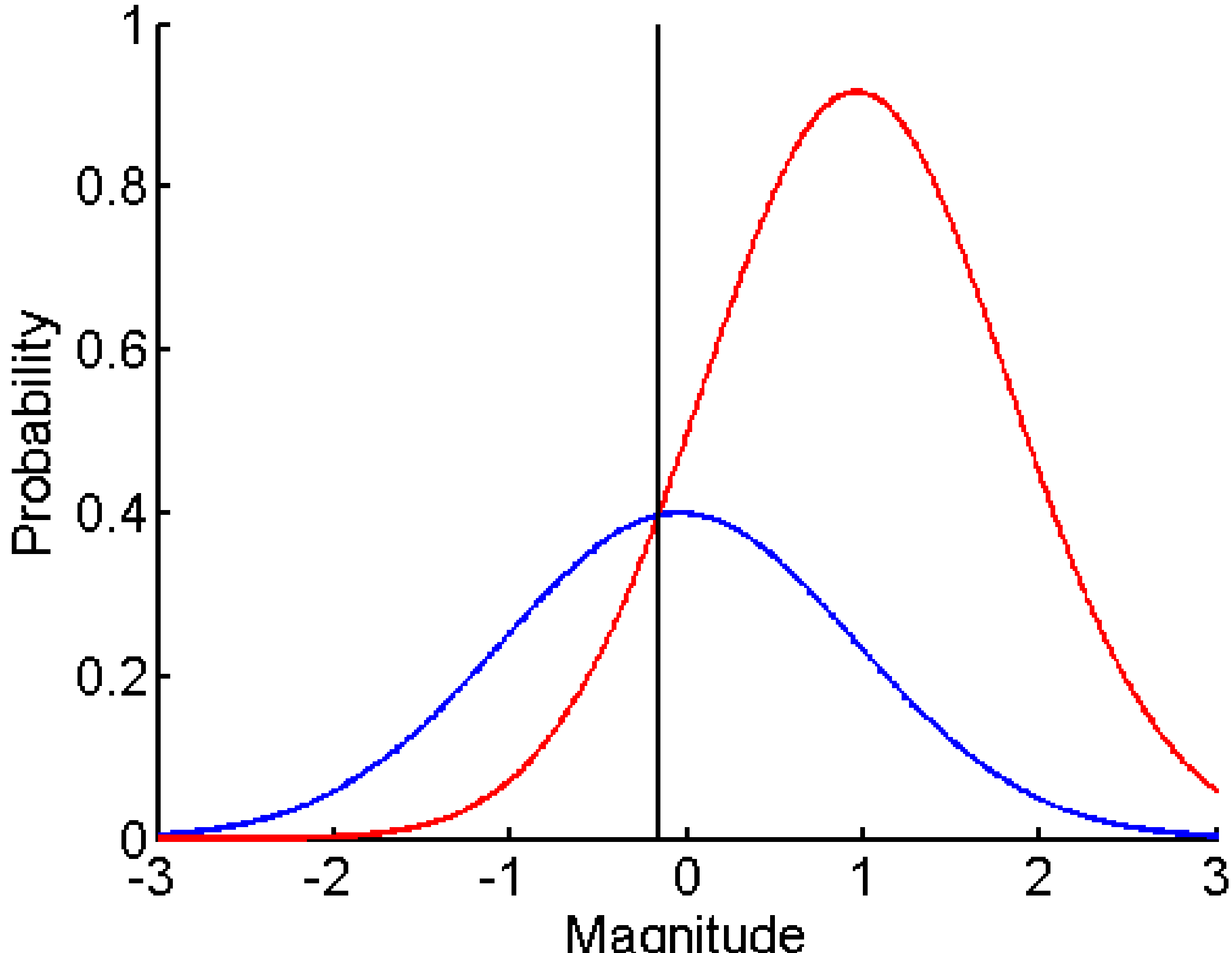
$$p\left(x \mid H_1\right) \gtrless p\left(x \mid H_2\right)$$

Minimum Error



$$P(H_1) = \frac{1}{3} \qquad P(H_2) = \frac{2}{3}$$





1
0.8
0.6
0.4
0.2
0
Probability
-3
-2
-1
0
1
2
3
Magnitude

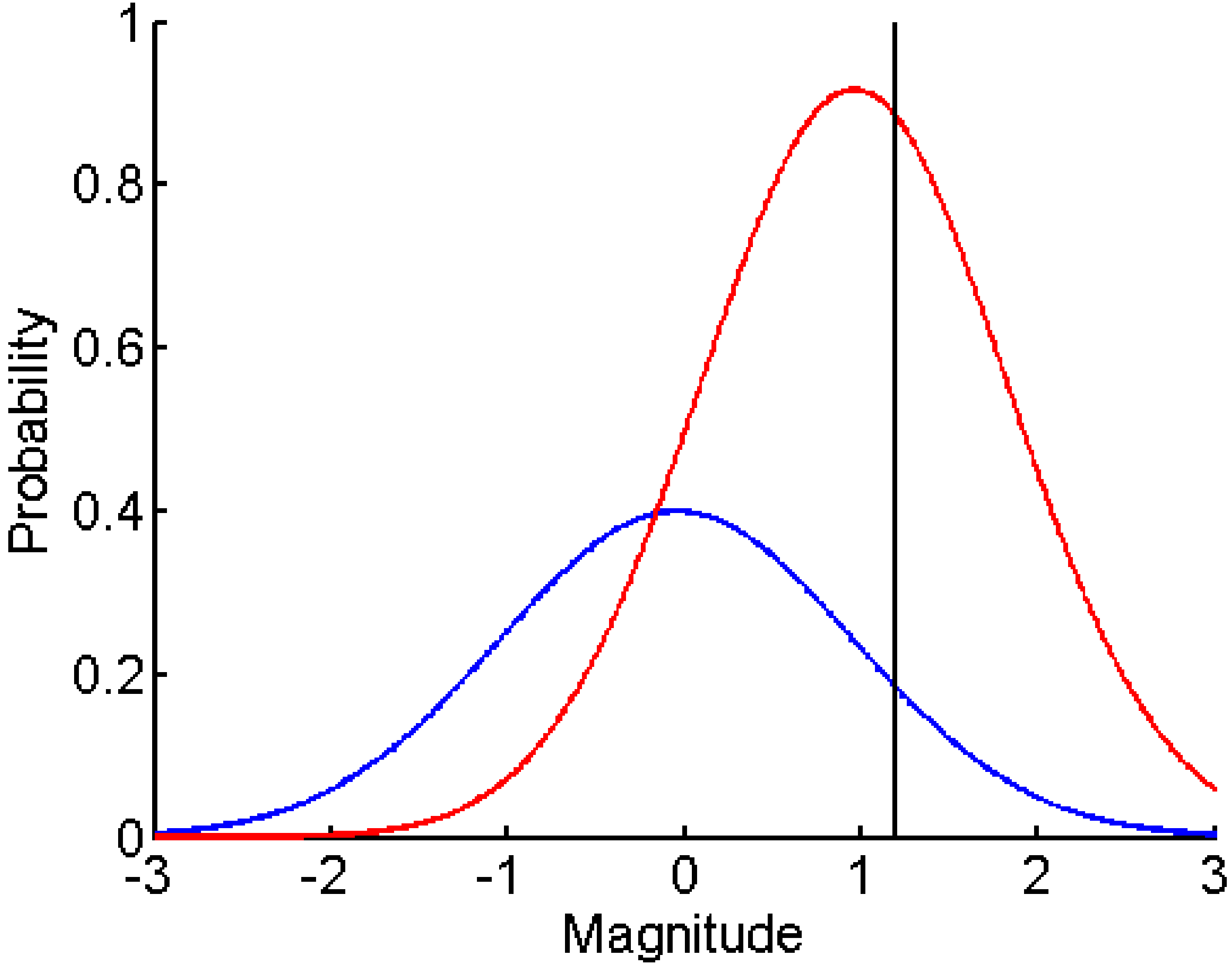
If all errors are not equally important, you can assign minimum cost rather than minimum error

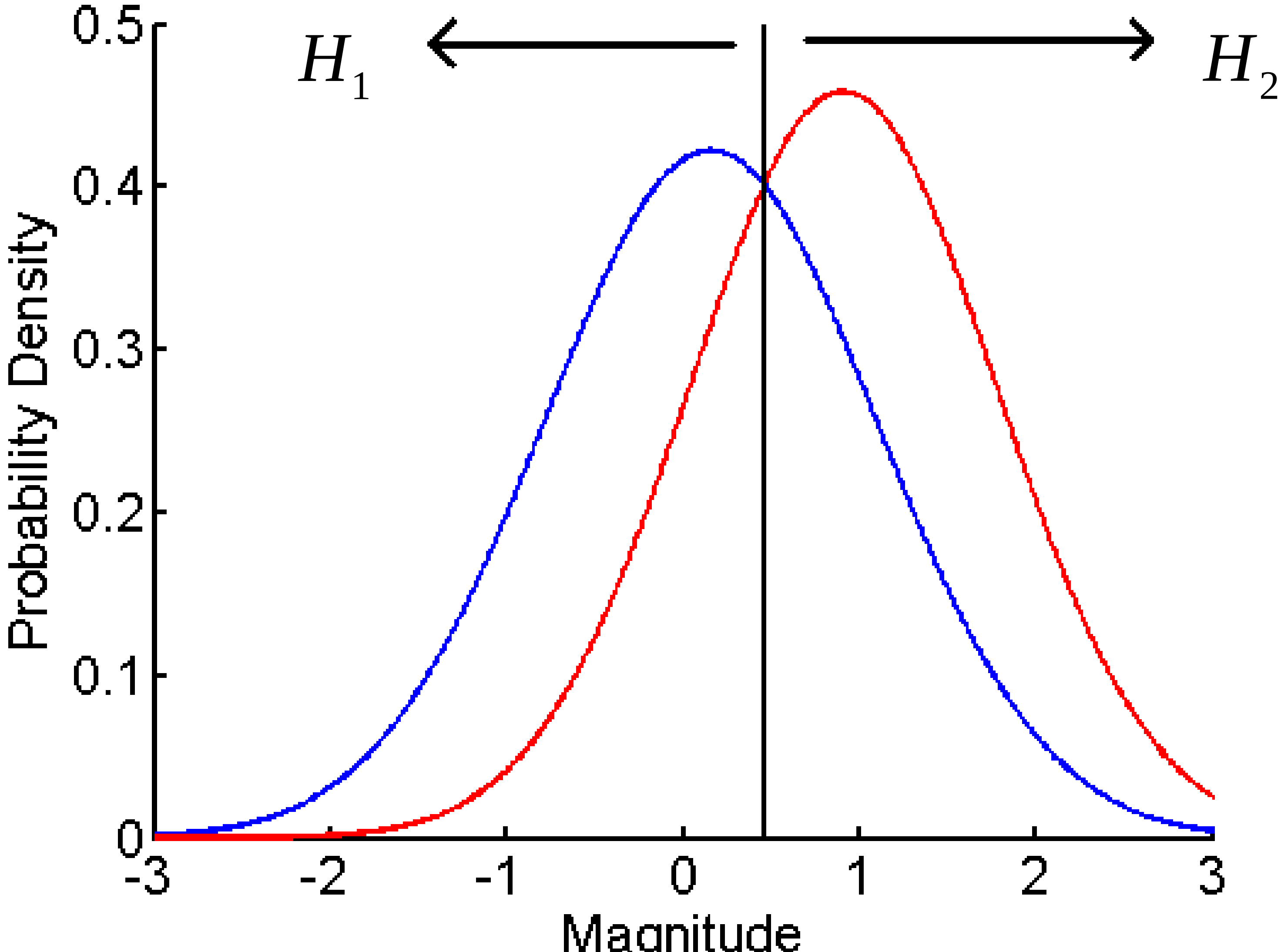
$$\boxed{\frac{p(x|H_1)}{p(x|H_2)} \underset{<}{>} \frac{(C_{12}-C_{22})}{(C_{21}-C_{11})}\frac{P(H_2)}{P(H_1)}}$$

If Gaussian distributions are assumed we can solve for a closed form classifier

$$\frac{(x-m_1)^2}{\sigma_1^2} - \frac{(x-m_2)^2}{\sigma_2^2} \underset{<}{>} \gamma$$

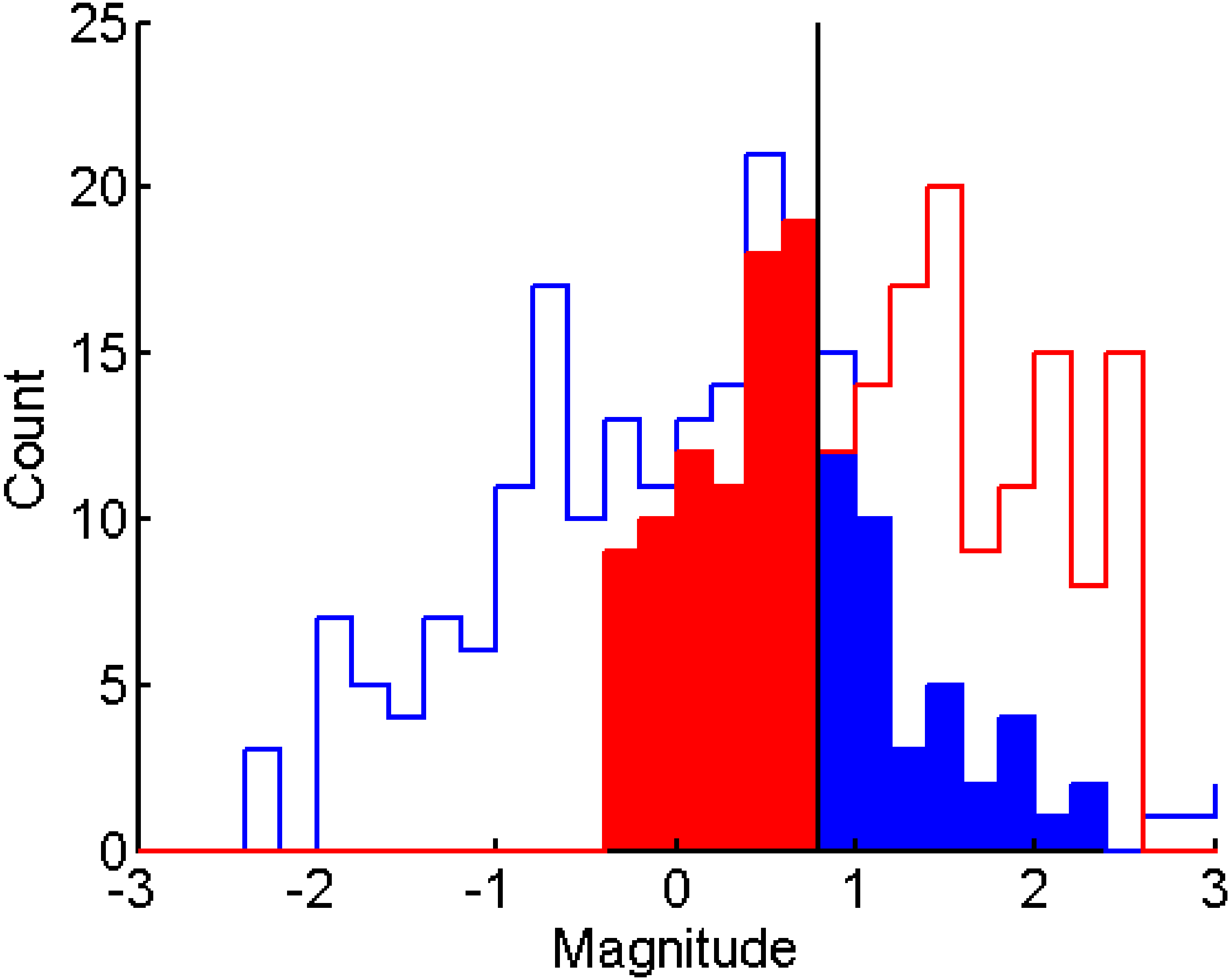
Minimum Cost



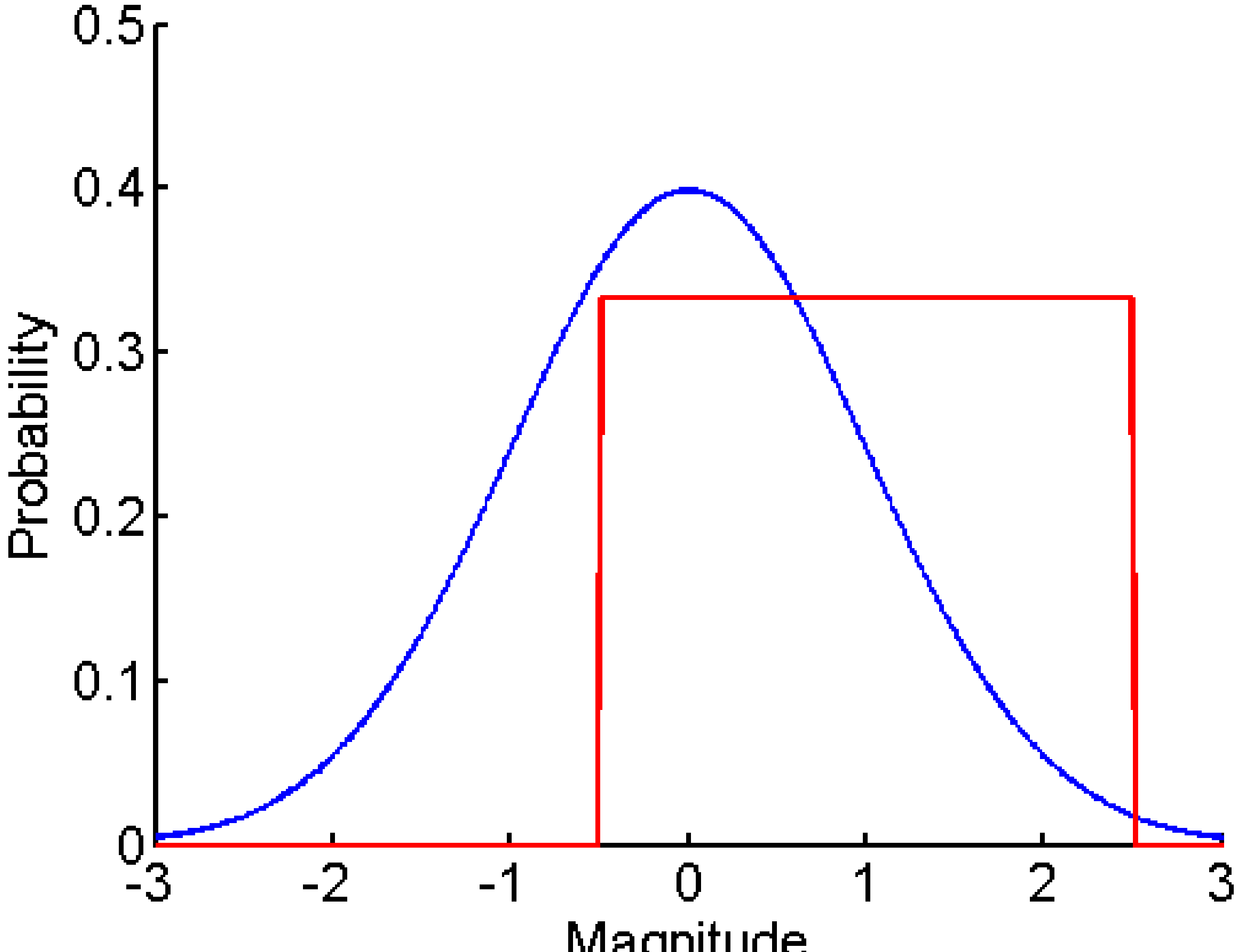


H_1
H_2
0.5
0.4
0.3
0.2
0.1
0
Probability Density
-3
-2
-1
0
1
2
3
Magnitude

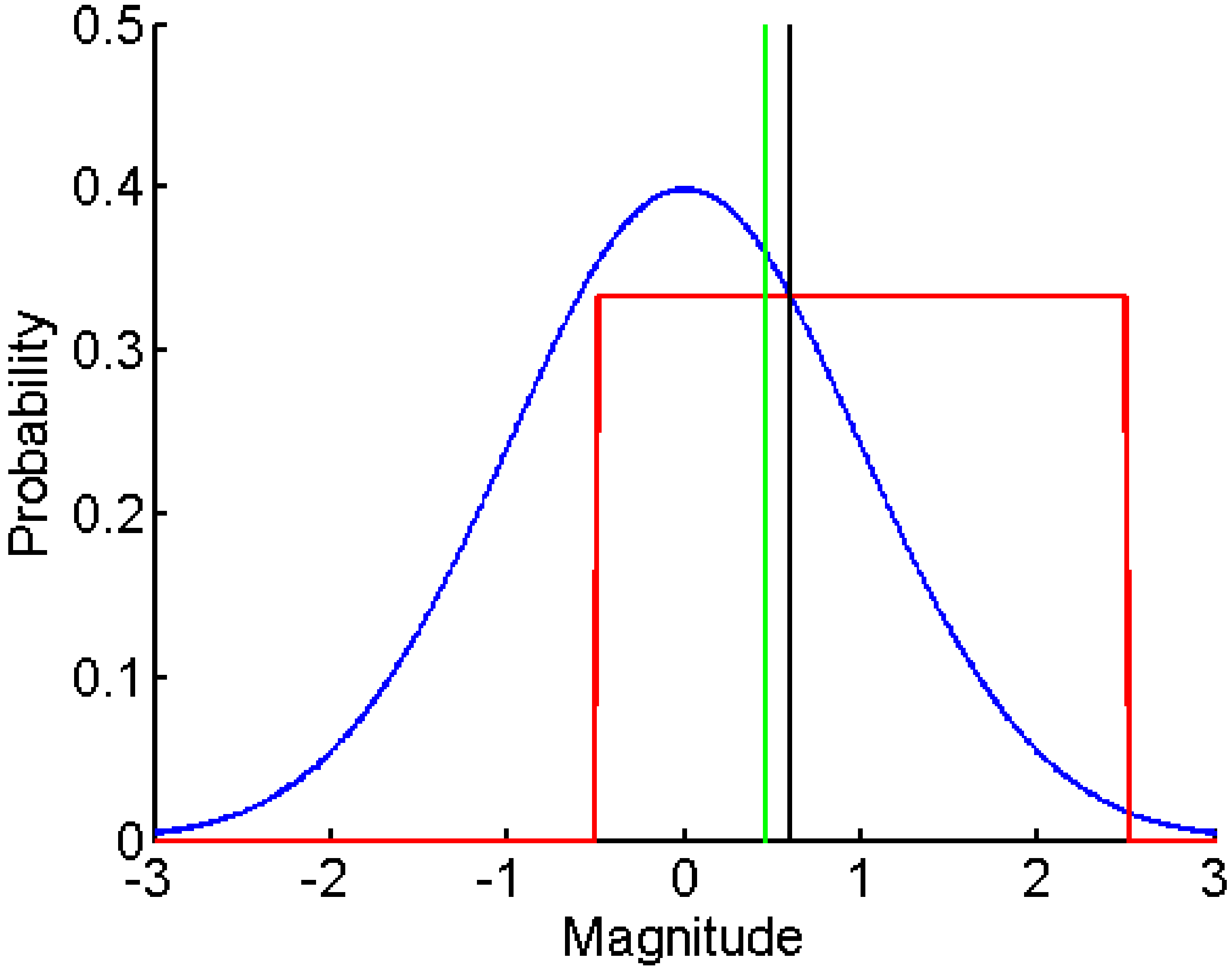
The classification error is higher than predicted!



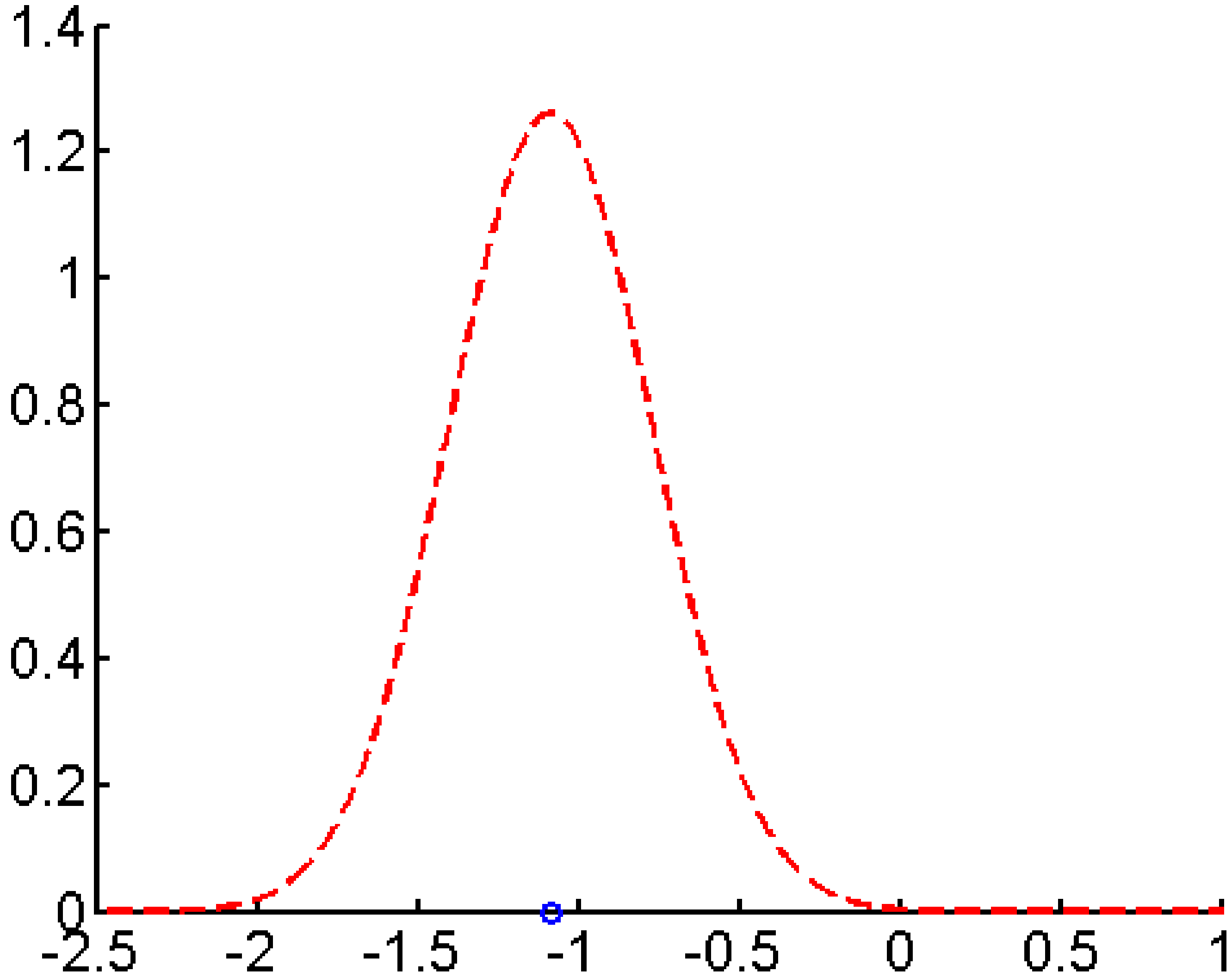
Because the underlying distributions are not Gaussian!



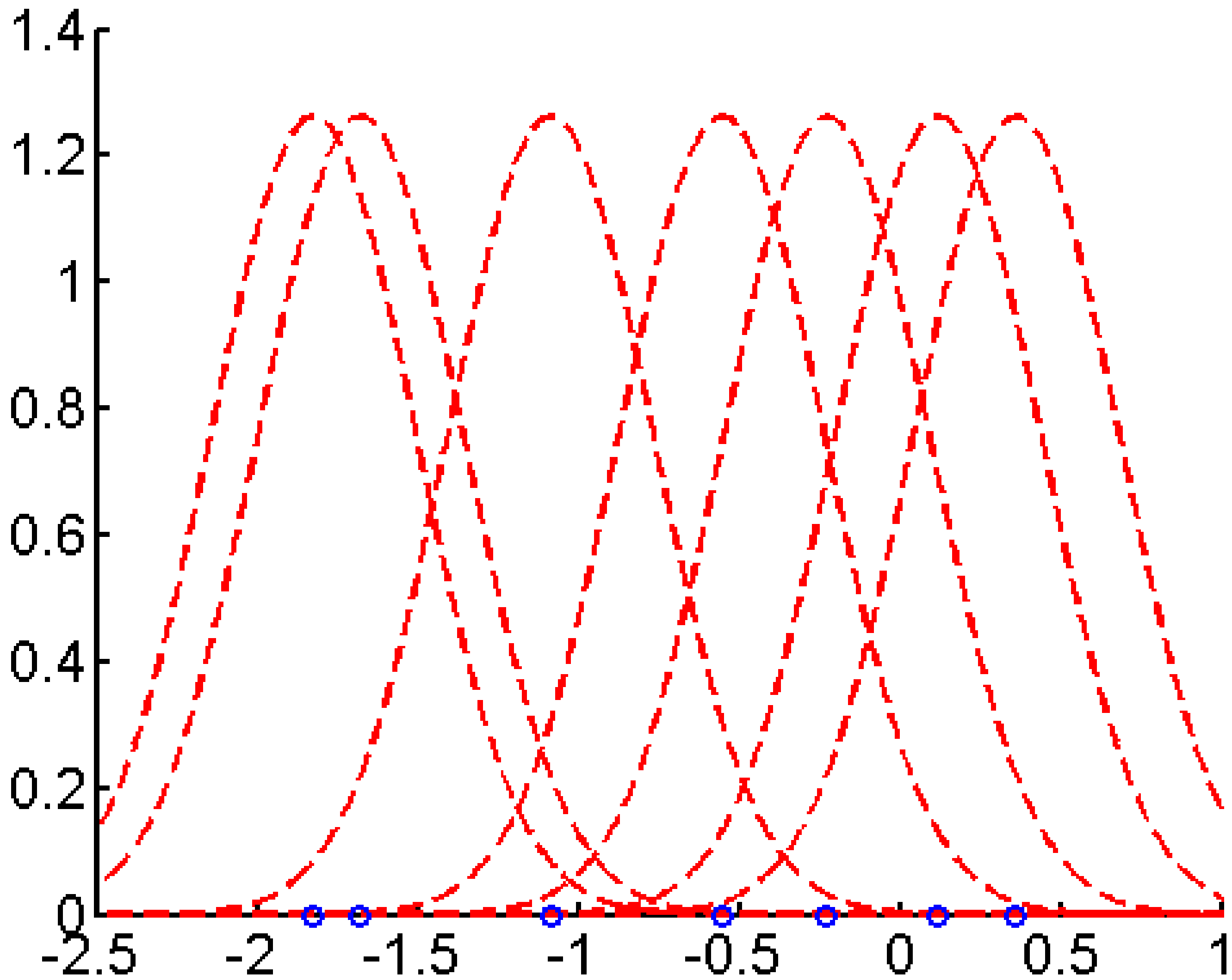
So we put the threshold in the wrong place



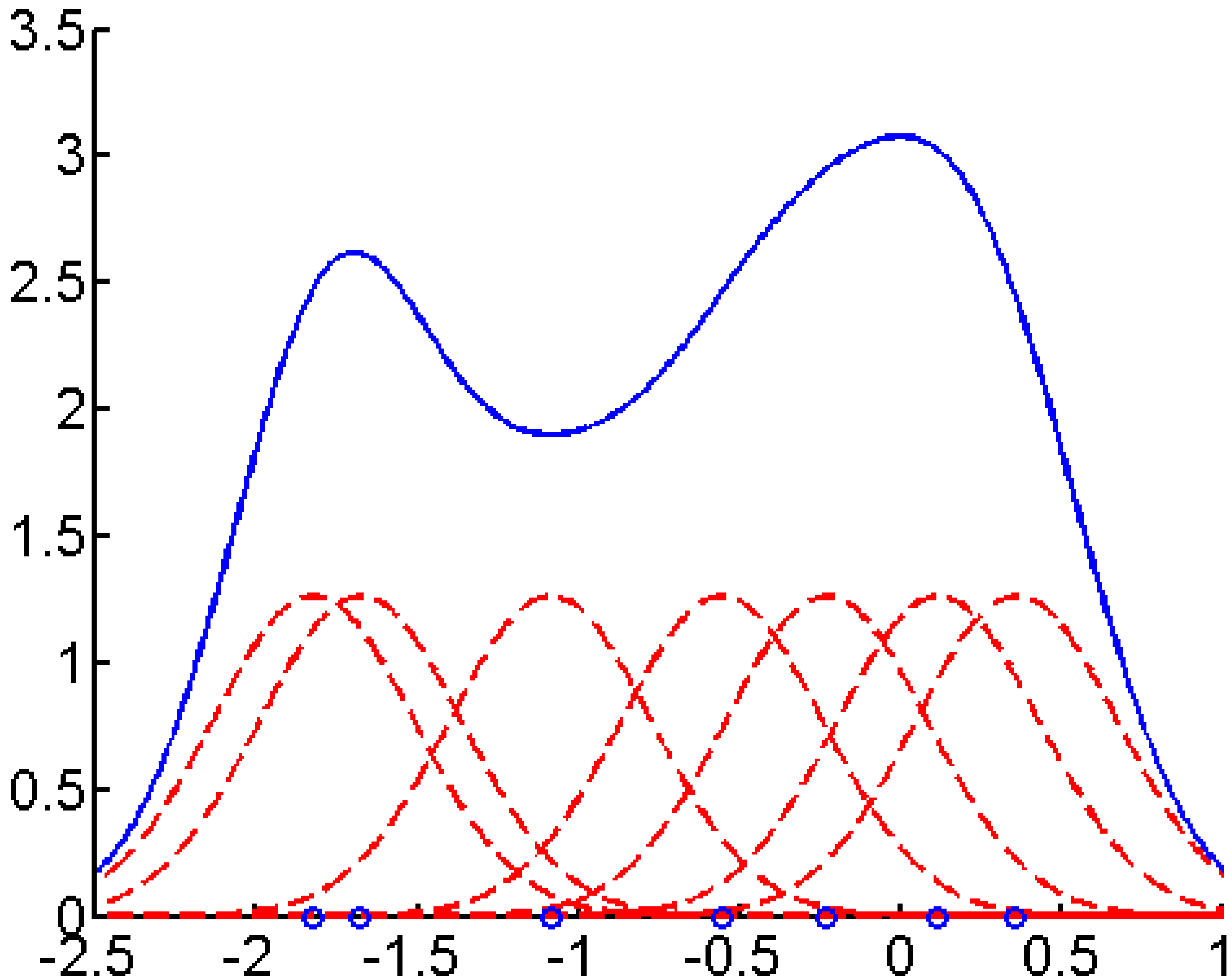
One way to estimate distributions: Kernel Density Estimates



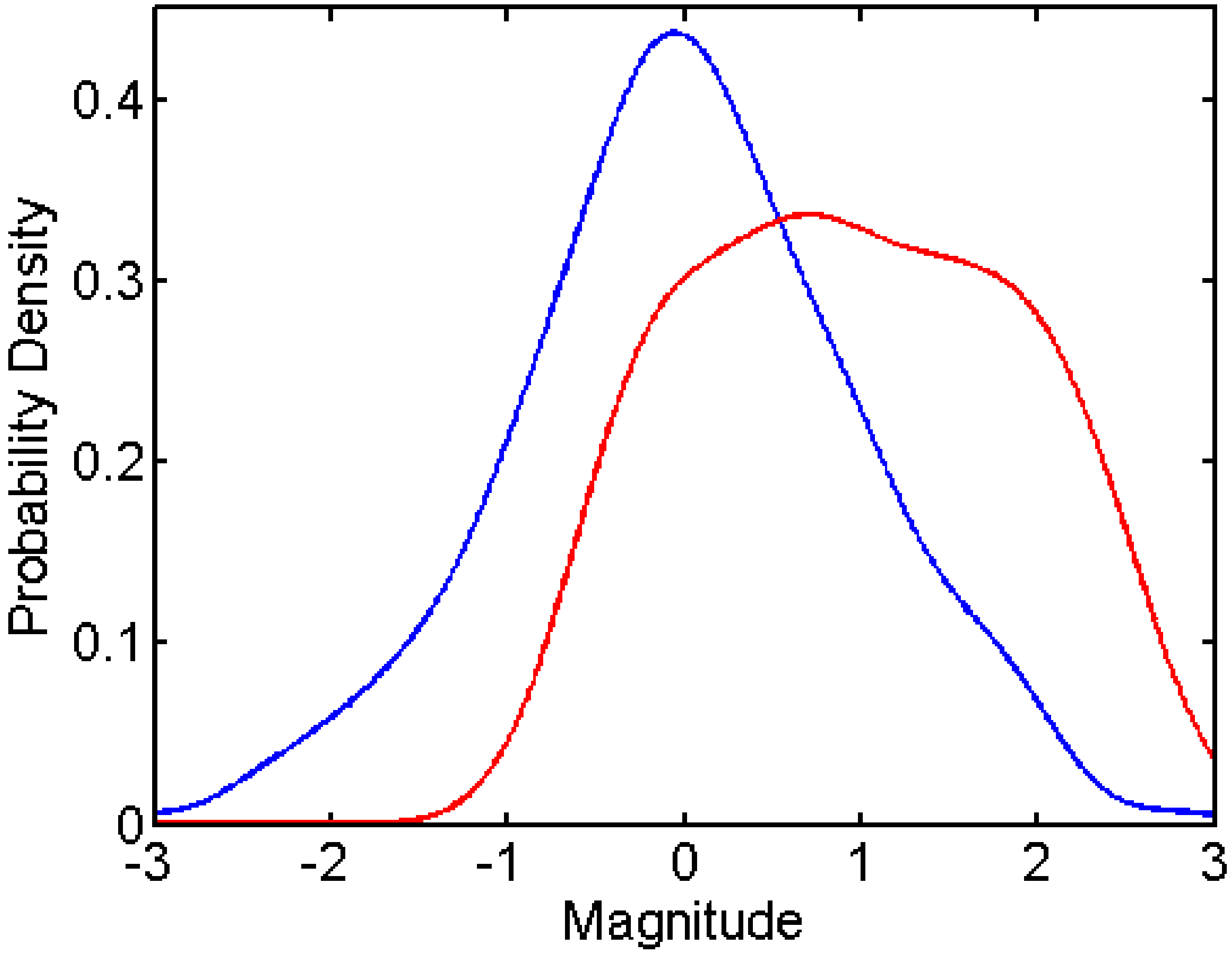
One way to estimate distributions: Kernel Density Estimates



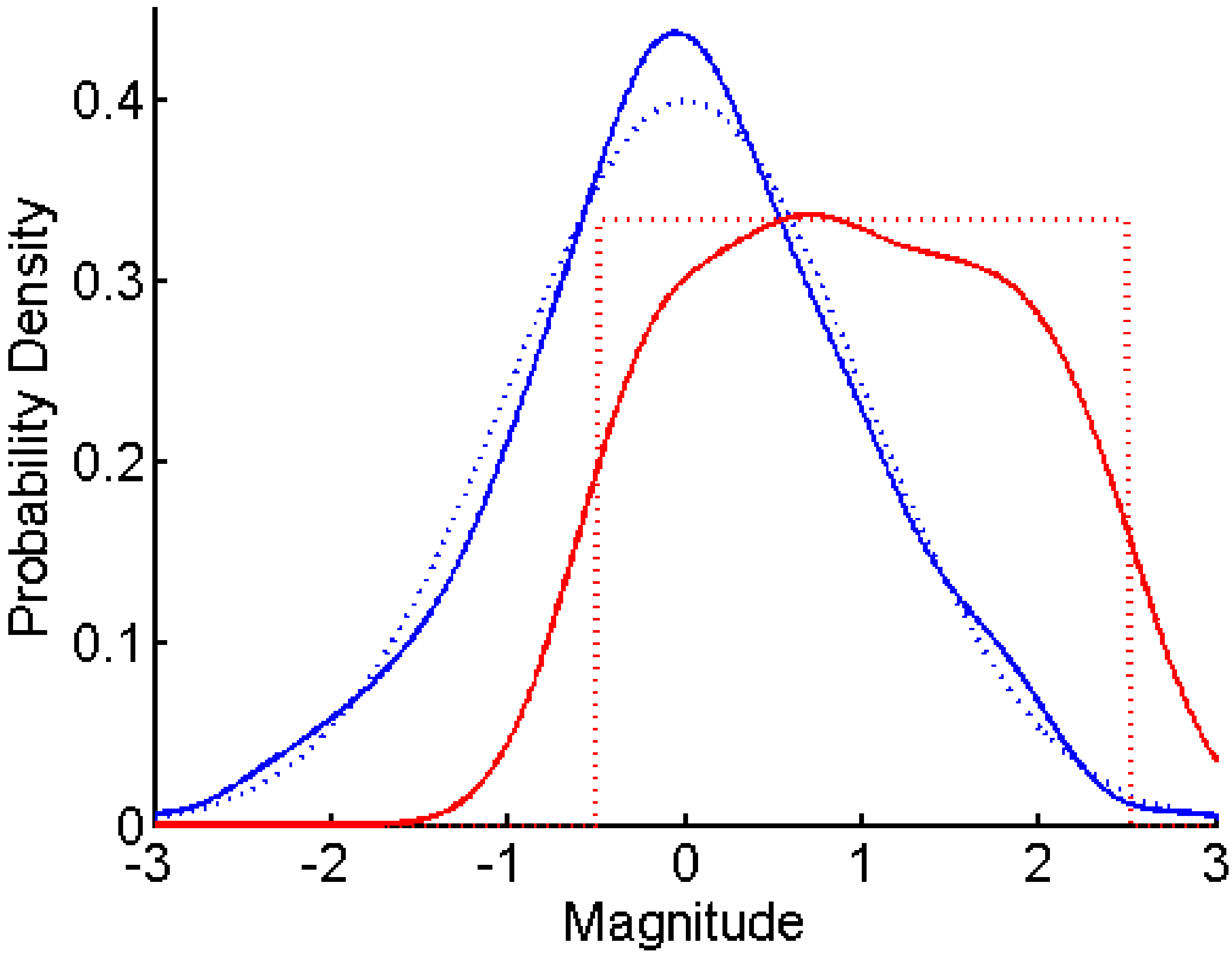
One way to estimate distributions:
Kernel Density Estimates



Kernel Density Estimates Results

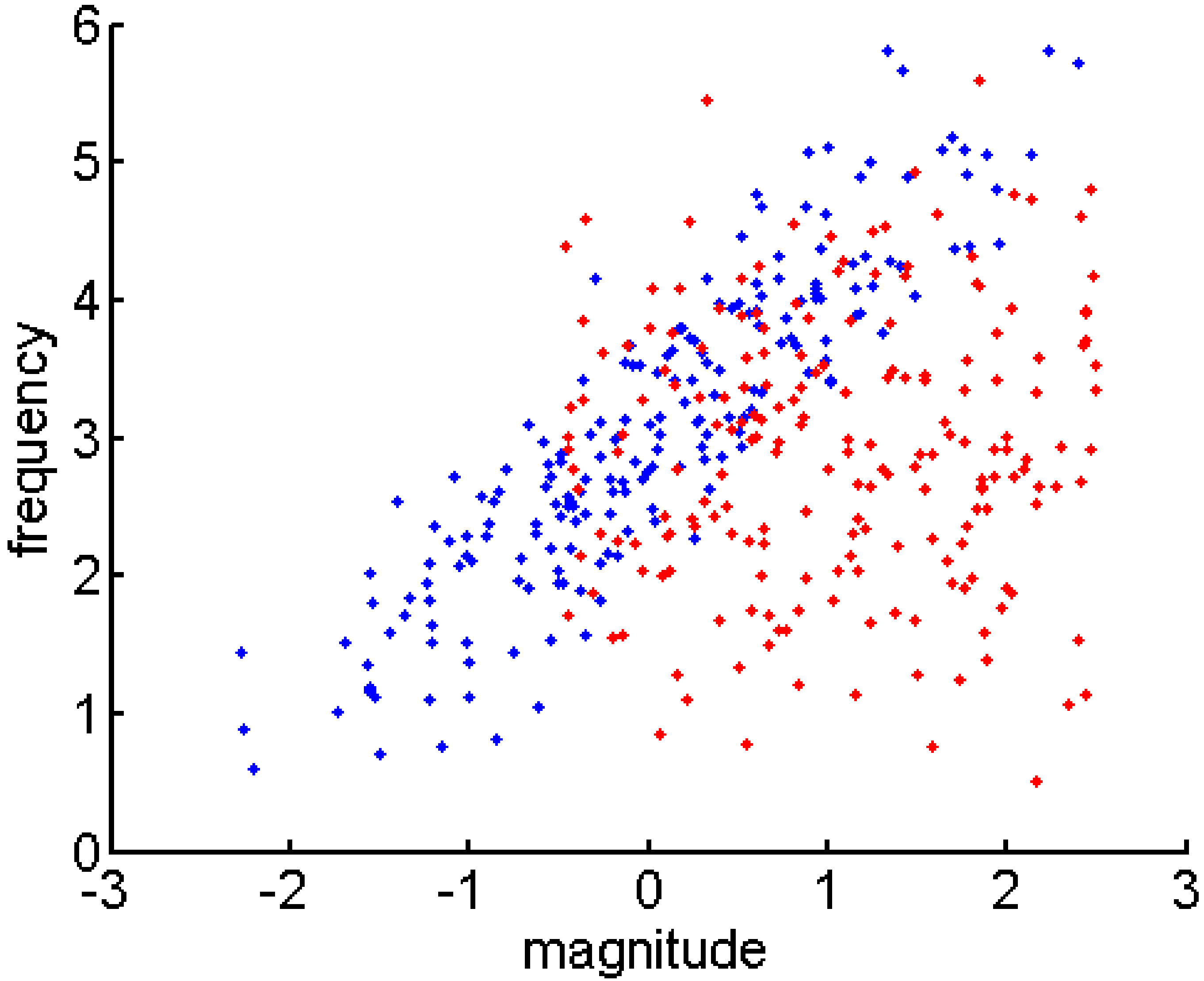


Kernel density method is optimal given infinite data…

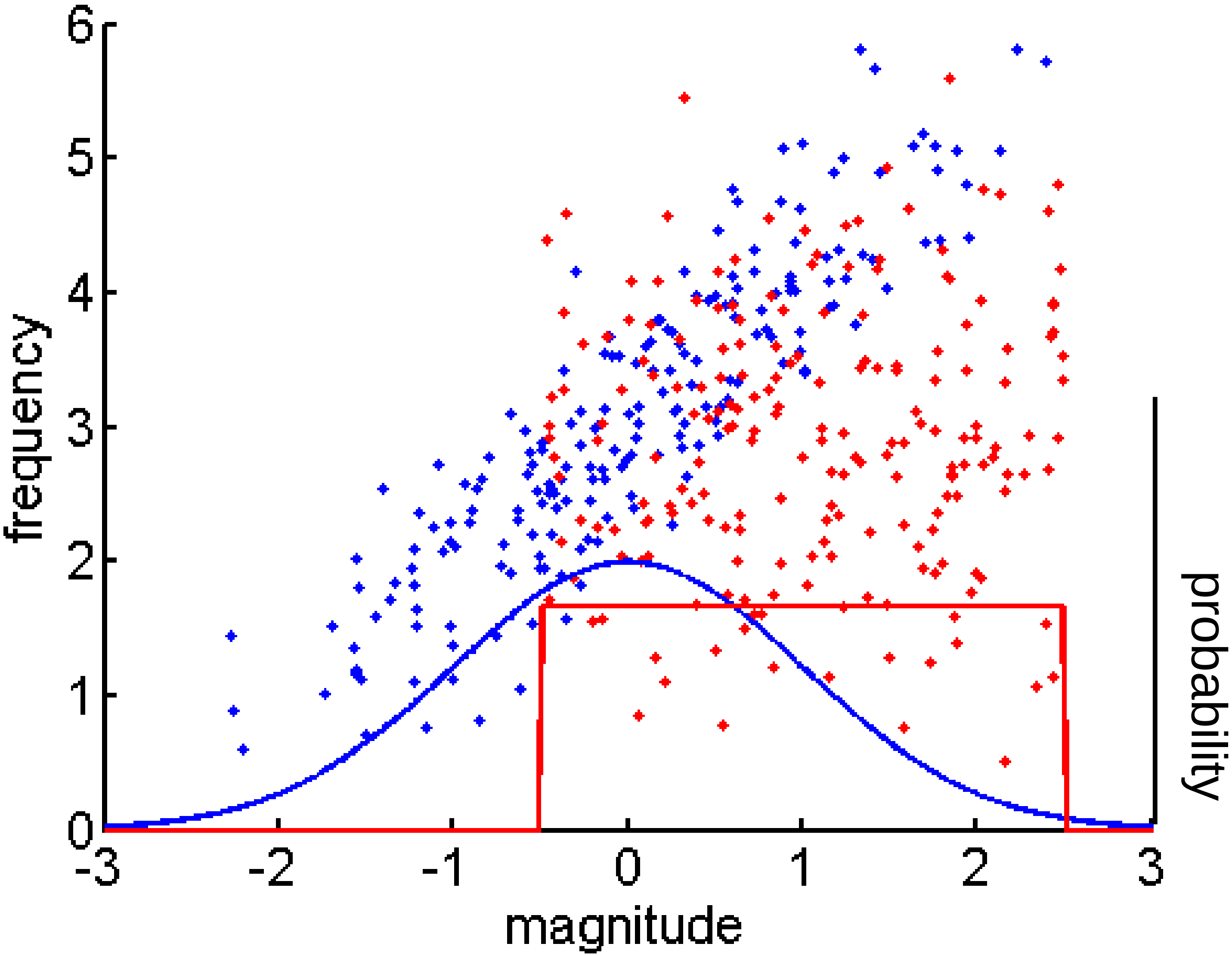


Why are we are able to classify marine mammal versus cargo ship much better than any of these methods?

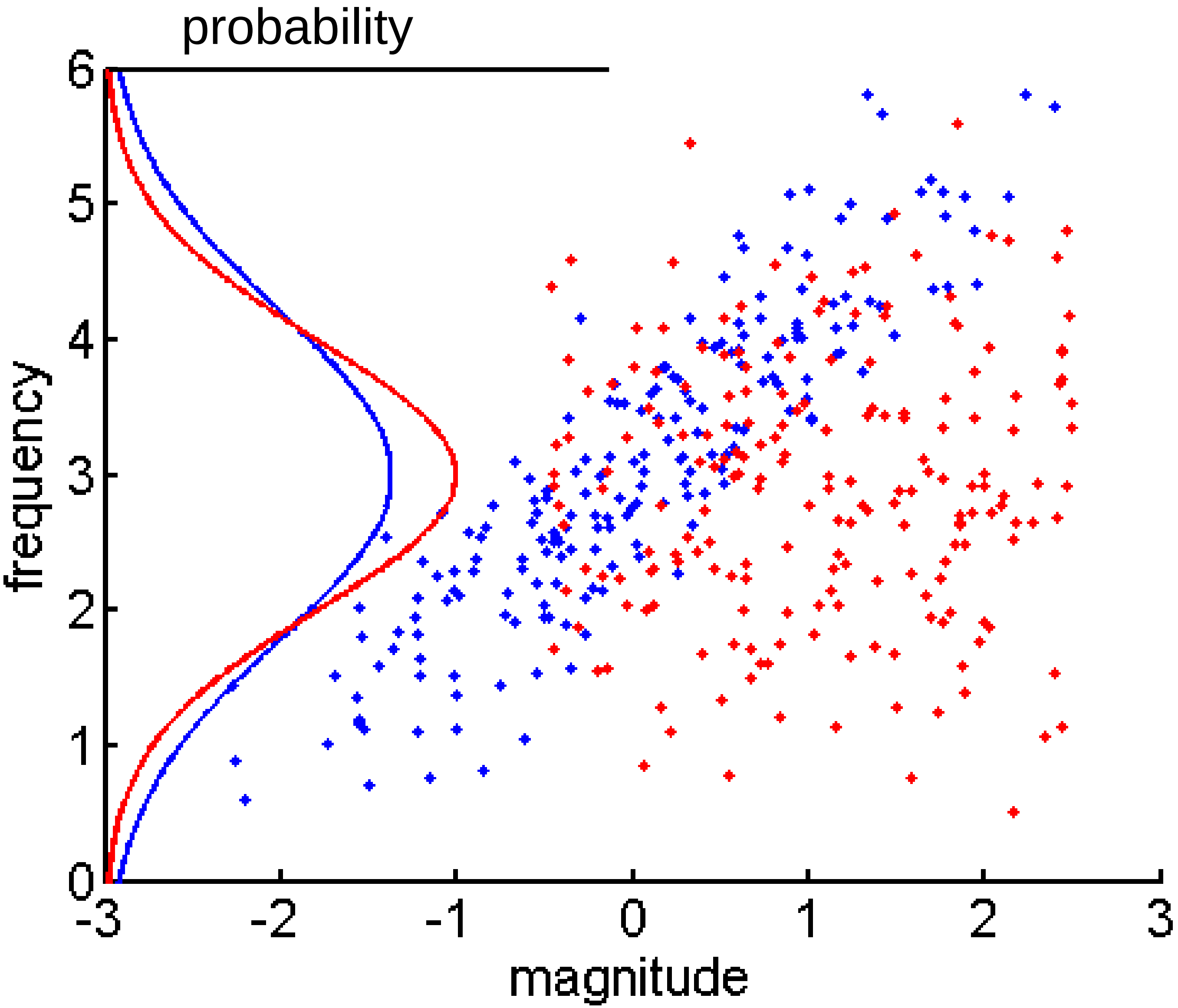
Because we don’t just listen to the amplitude!



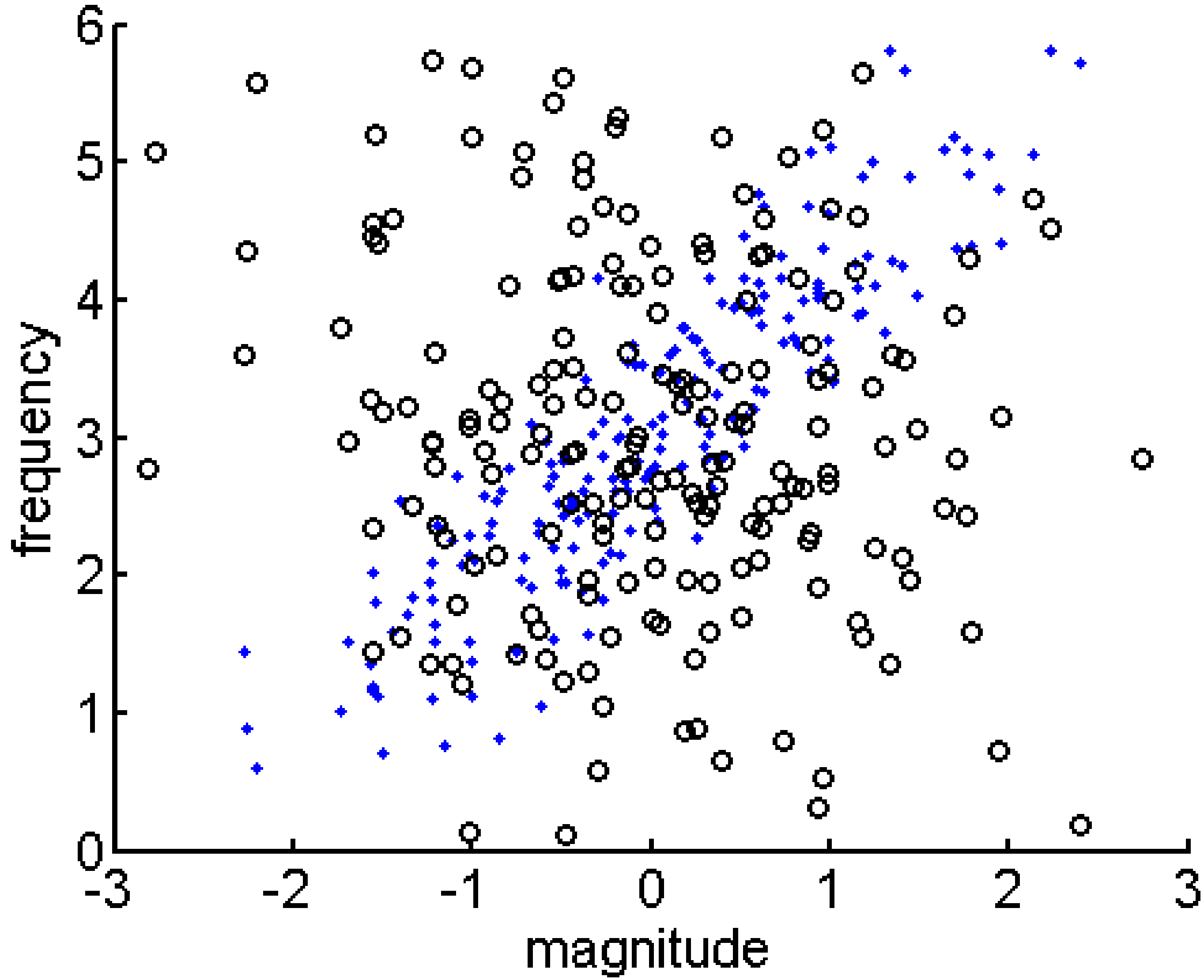
6
5
4
3
2
1
0
-3
-2
-1
0
1
2
3
frequency
magnitude



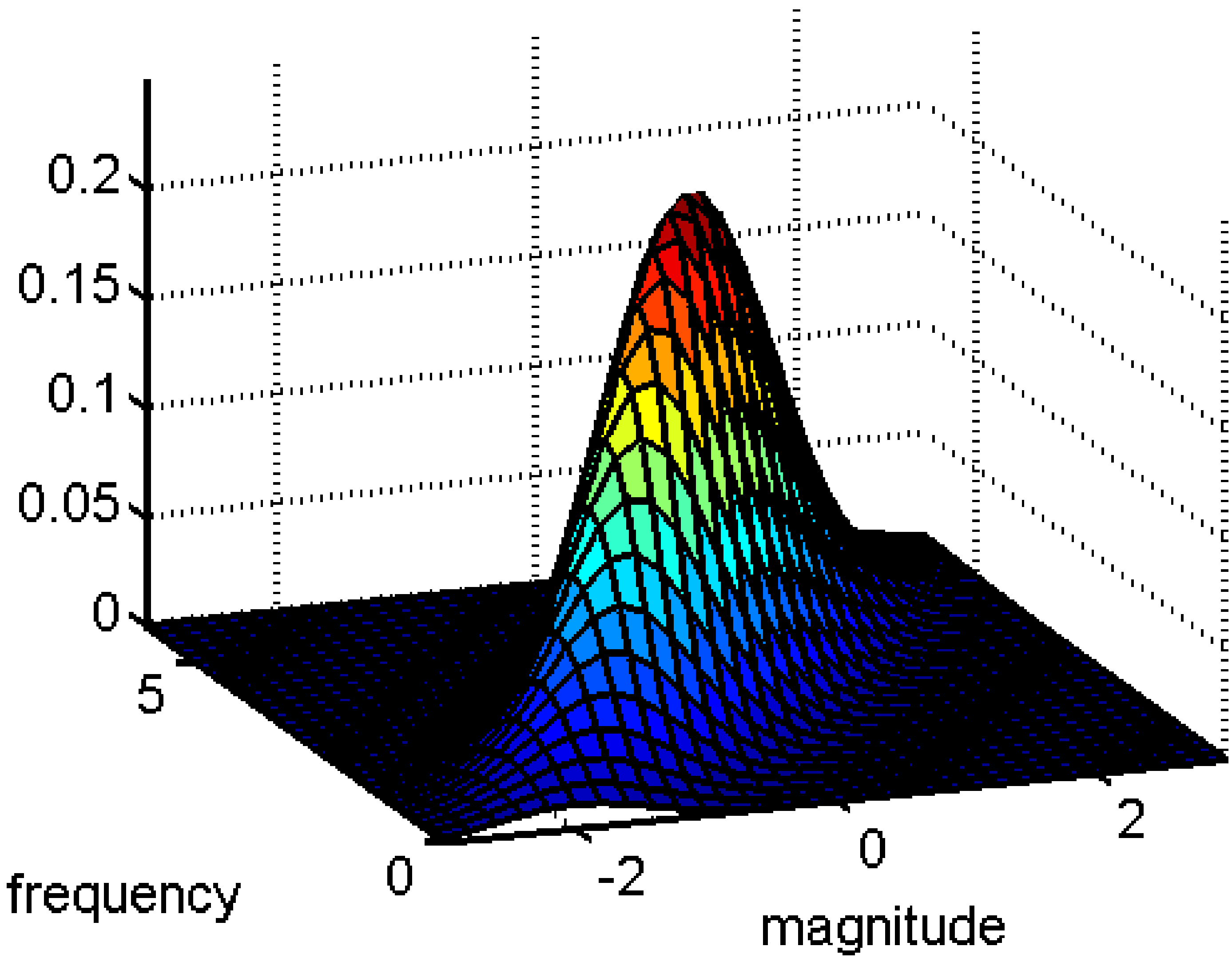
frequency
6
5
4
3
2
1
0
-3
-2
-1
0
1
2
3
magnitude
probability


probability
frequency
magnitude
0
1
2
3
4
5
6
-3
-2
-1
0
1
2
3

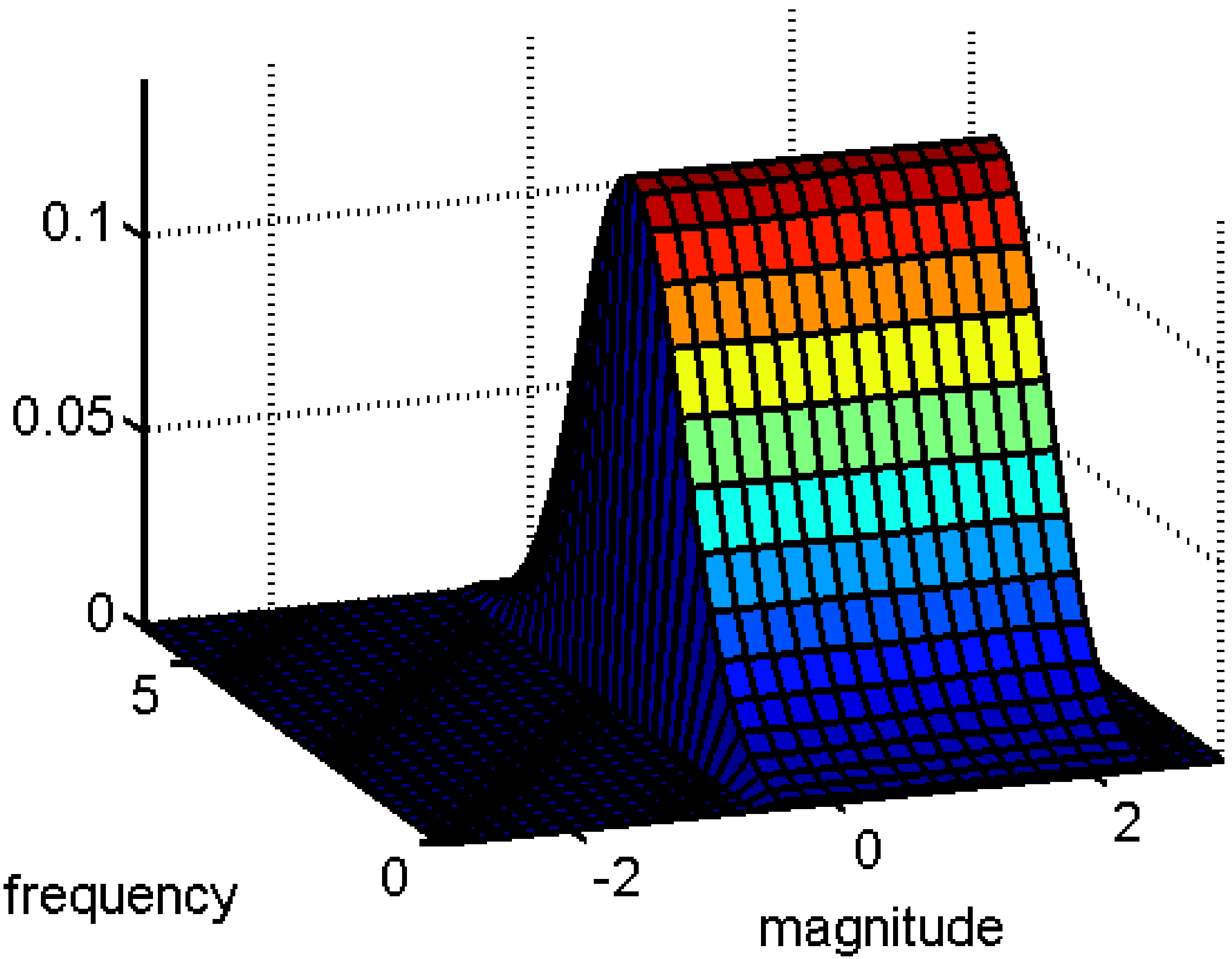
$$p(x, y) \neq p(x)\,p(y)$$

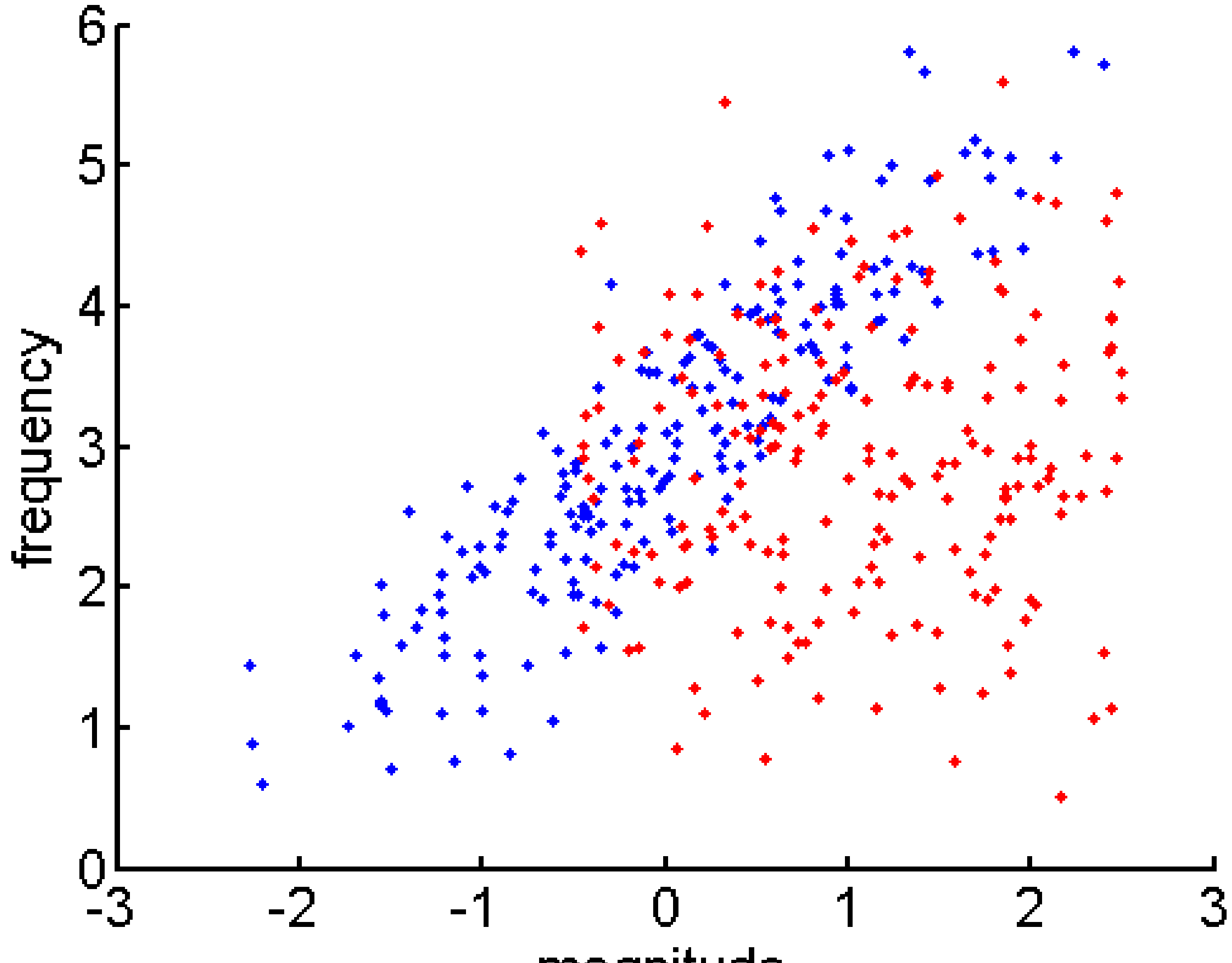


Hypothesis 1: Marine Mammal



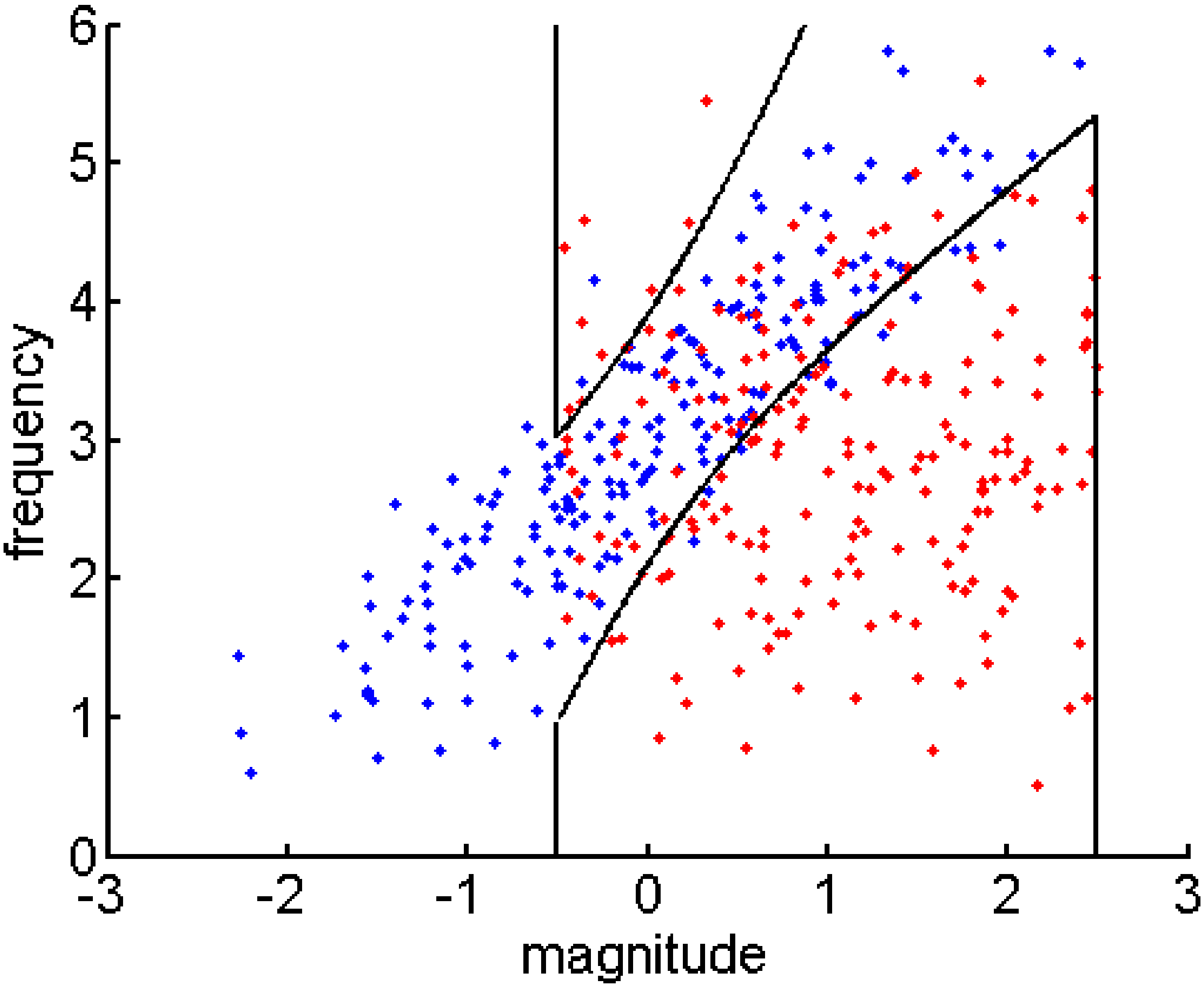
Hypothesis 2: Cargo Ship





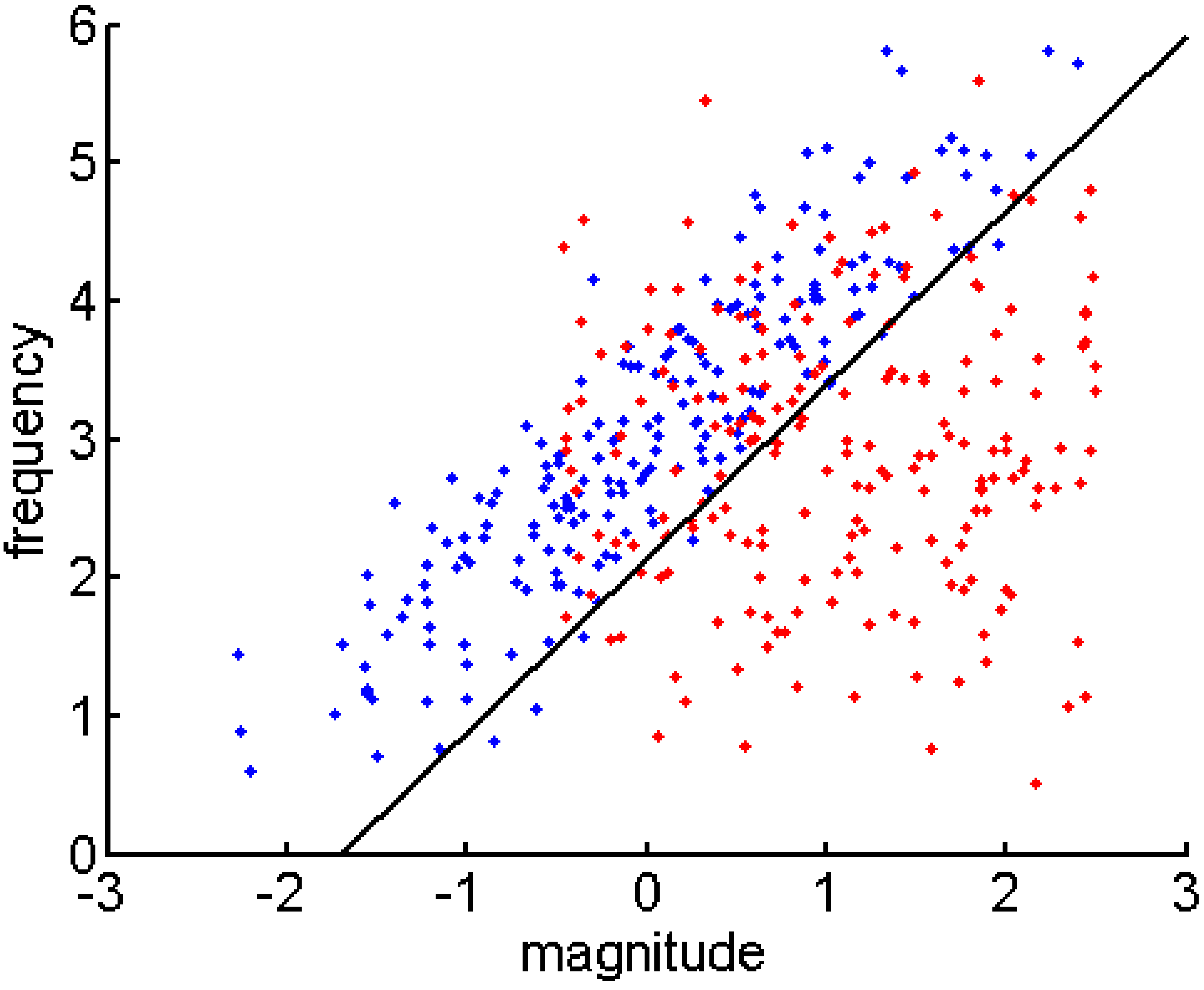
frequency
magnitude
6
5
4
3
2
1
0
-3
-2
-1
0
1
2
3

Likelihood Ratio Test

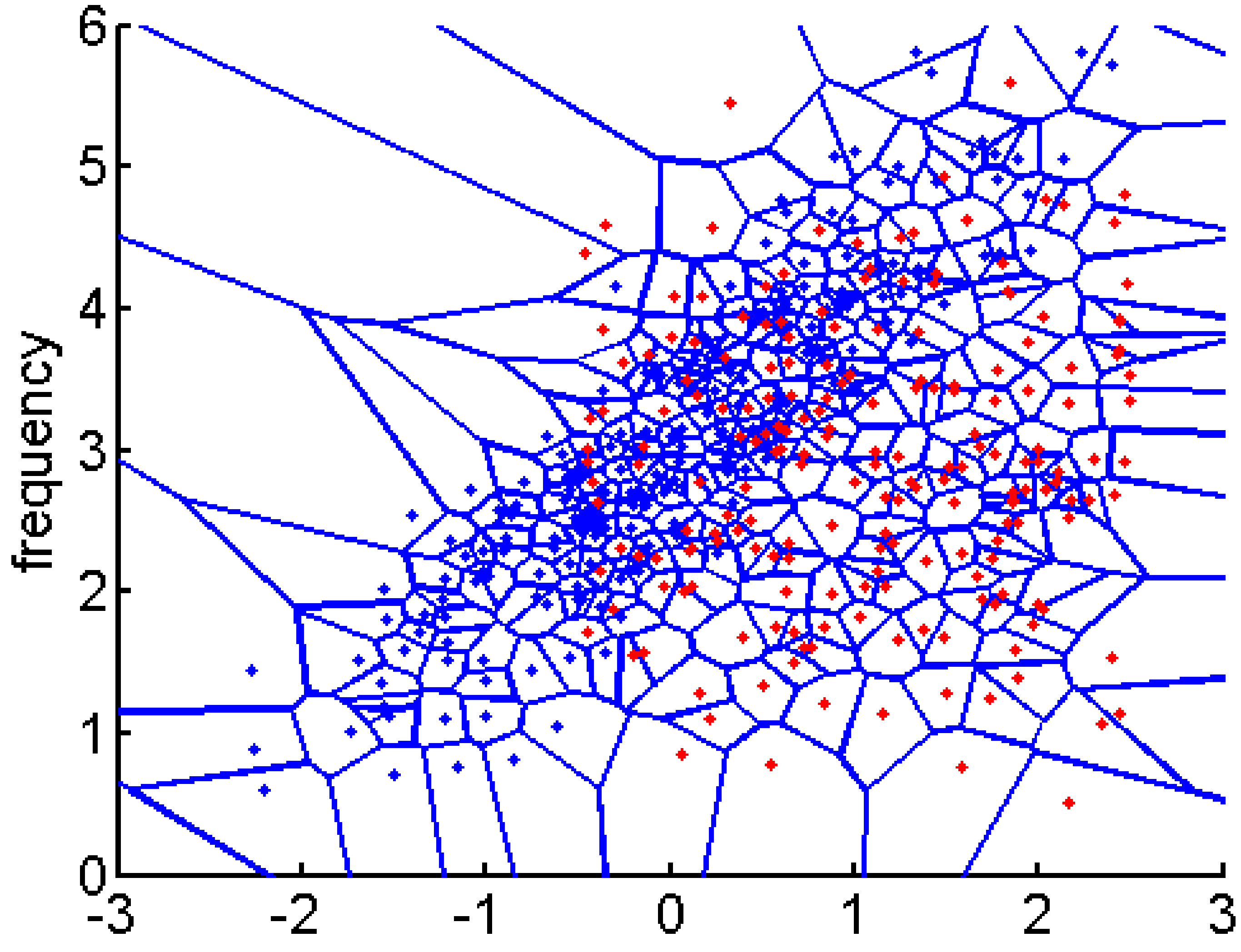


Minimum Error Linear Classifier

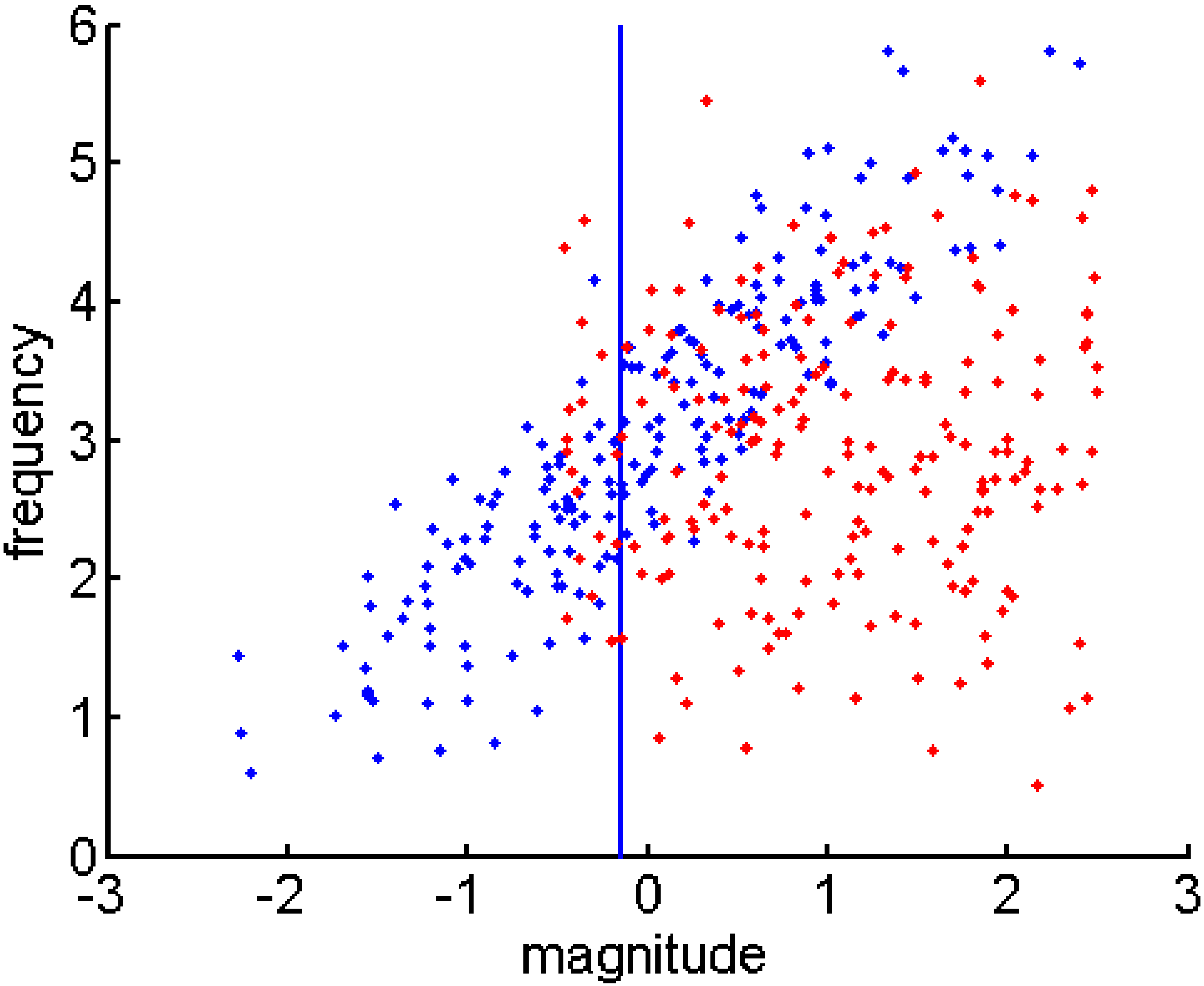
(SVM, Neural Net, etc)



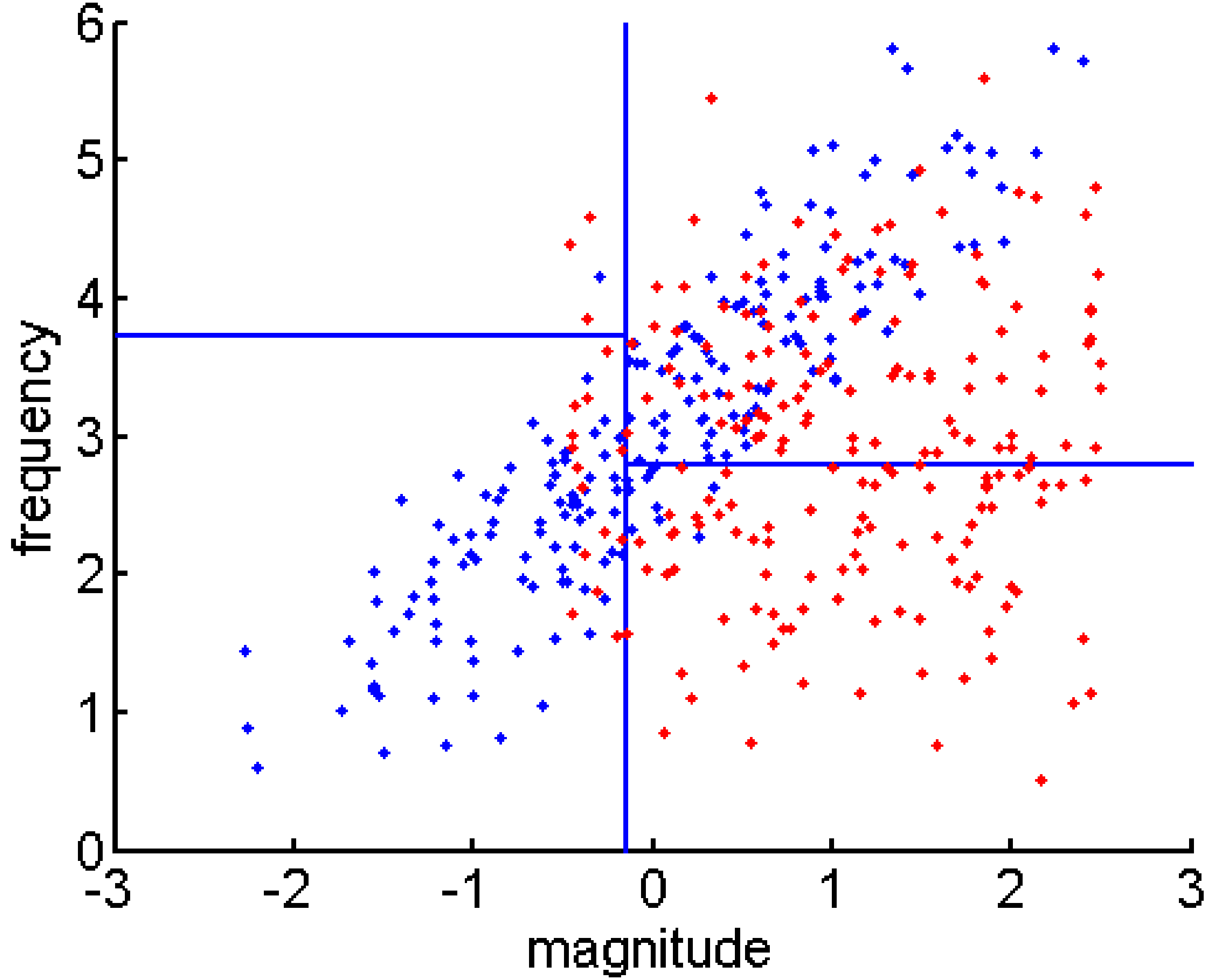
Nearest Neighbor



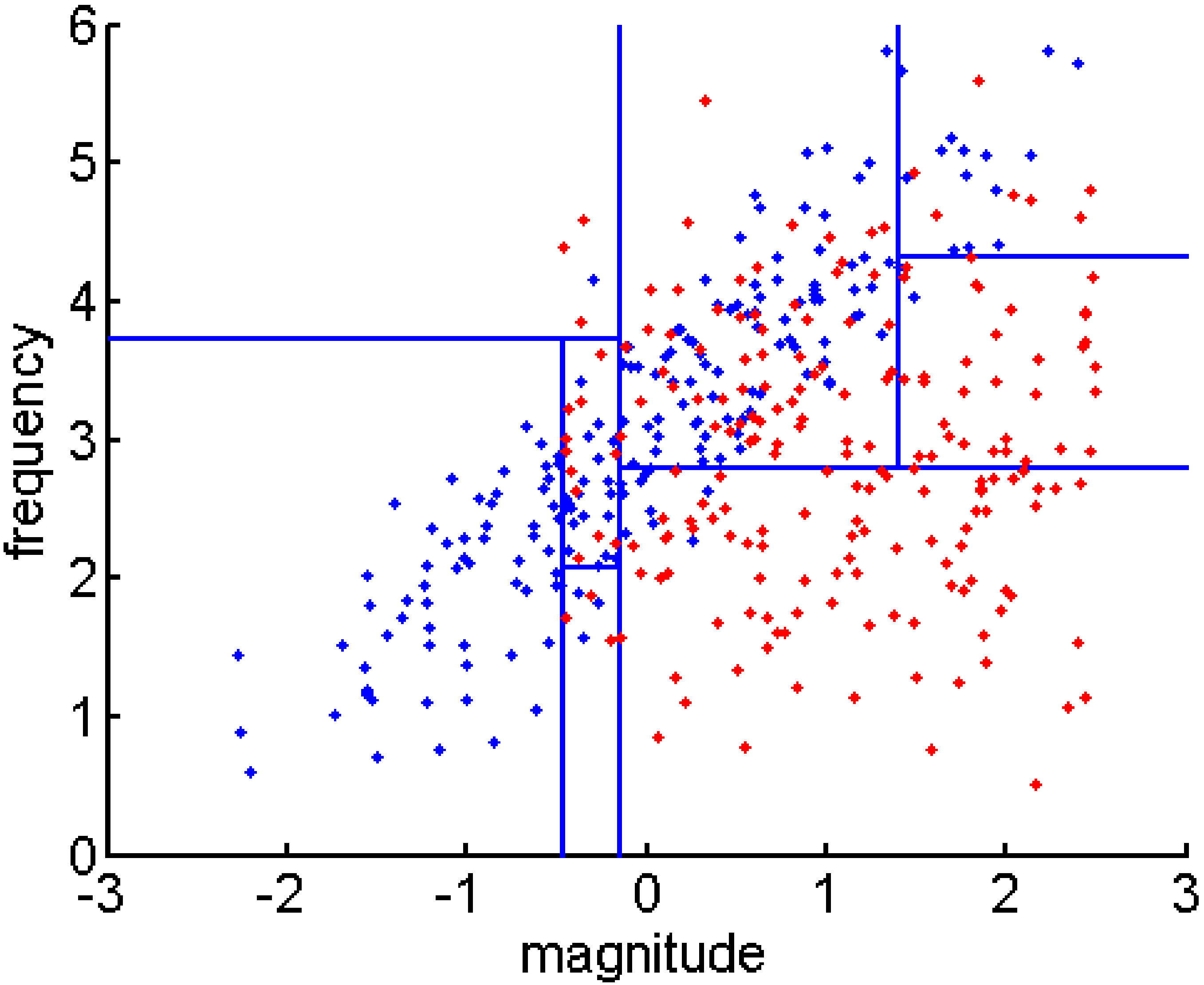
Decision Trees



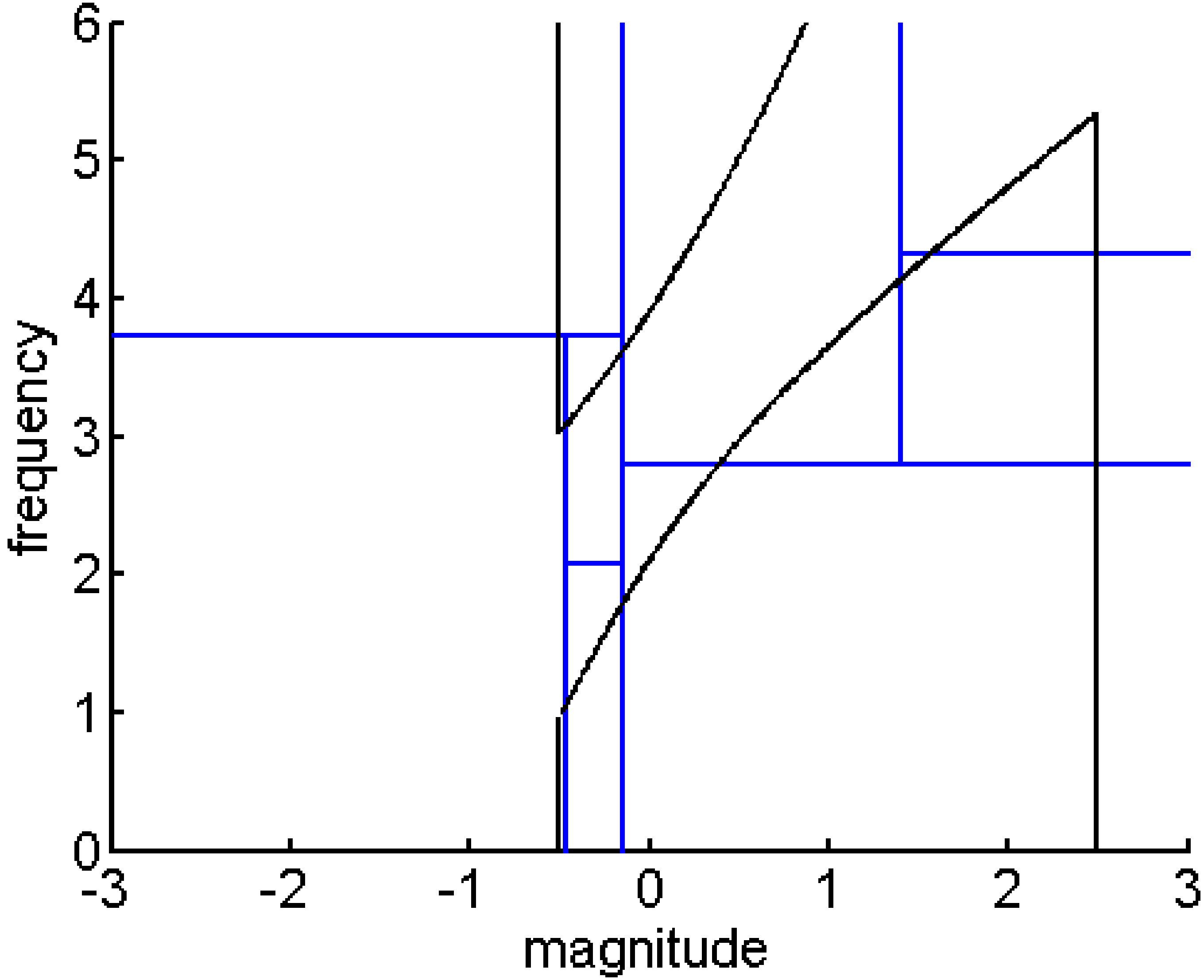
Decision Trees



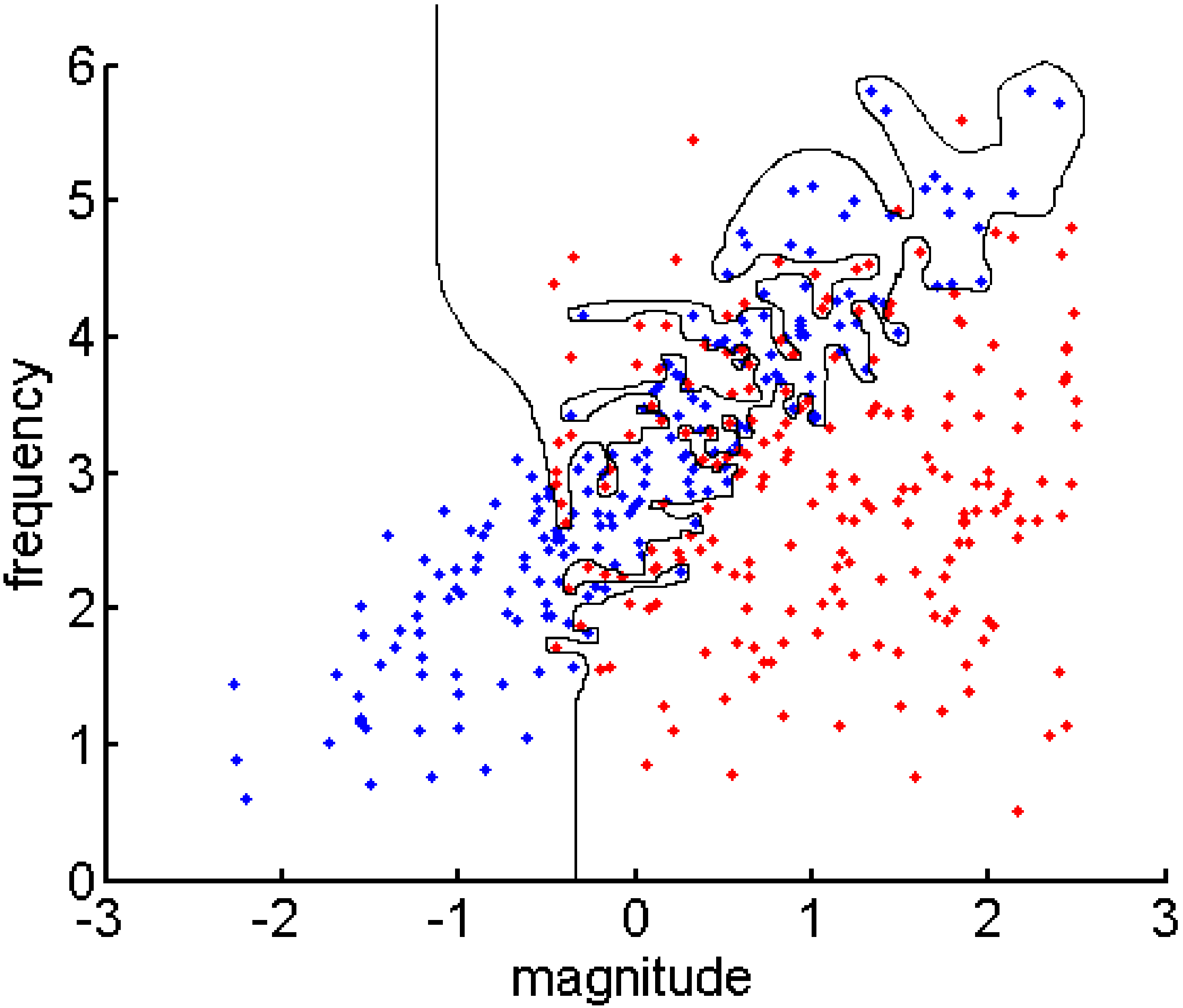
Decision Trees



Decision Trees



Over fitting - poor generalization



Implementation Issues:

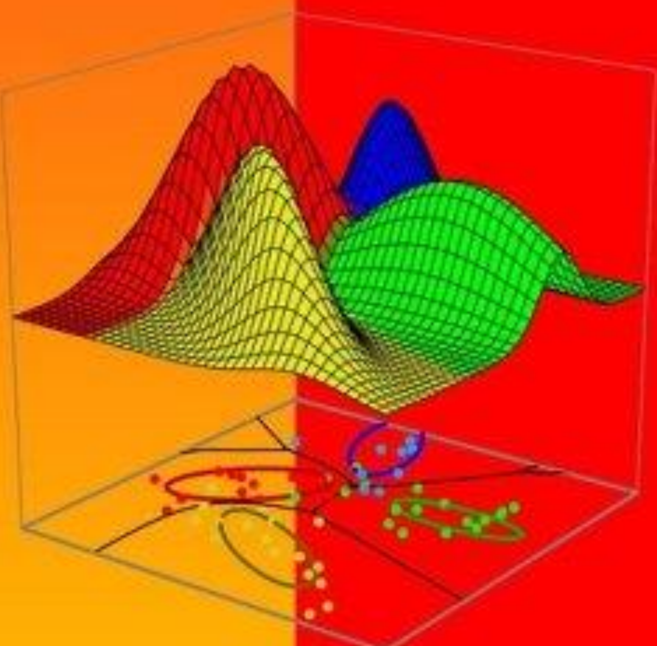
- Dimensionality & Computational Complexity
- Need for labeled data
- Choosing a feature set
  - Best features come from physics or domain specific knowledge
  - There are automated tools
  - All these methods really do is divide up feature space

Conclusions

- Many powerful techniques exist for automatic pattern classification
- The math / jargon should not scare you off!
- Many potential applications in acoustic signal processing

Richard O. Duda
Peter E. Hart
David G. Stork

Pattern Classification



Second Edition