



NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

Data Science

Classification Tree

Sourav SEN GUPTA, Lecturer

School of Computer Science and Engg.
Nanyang Technological University



Second Motivation for Data Science

CLASSIFICATION IN BUSINESS



The Science Common Problems

Prediction : Categorical

Is it Type A
or Type B?

What are the chances
that Sourav will default
a car purchase loan?

Will DBS allow one?



<https://www.youtube.com/watch?v=0XyV91VYrDs>



NANYANG TECHNOLOGICAL UNIVERSITY | SINGAPORE

The Science Common Solutions

Prediction : Numeric

Regression

**Prediction : Categorical
Classification**

Detection : Anomalies

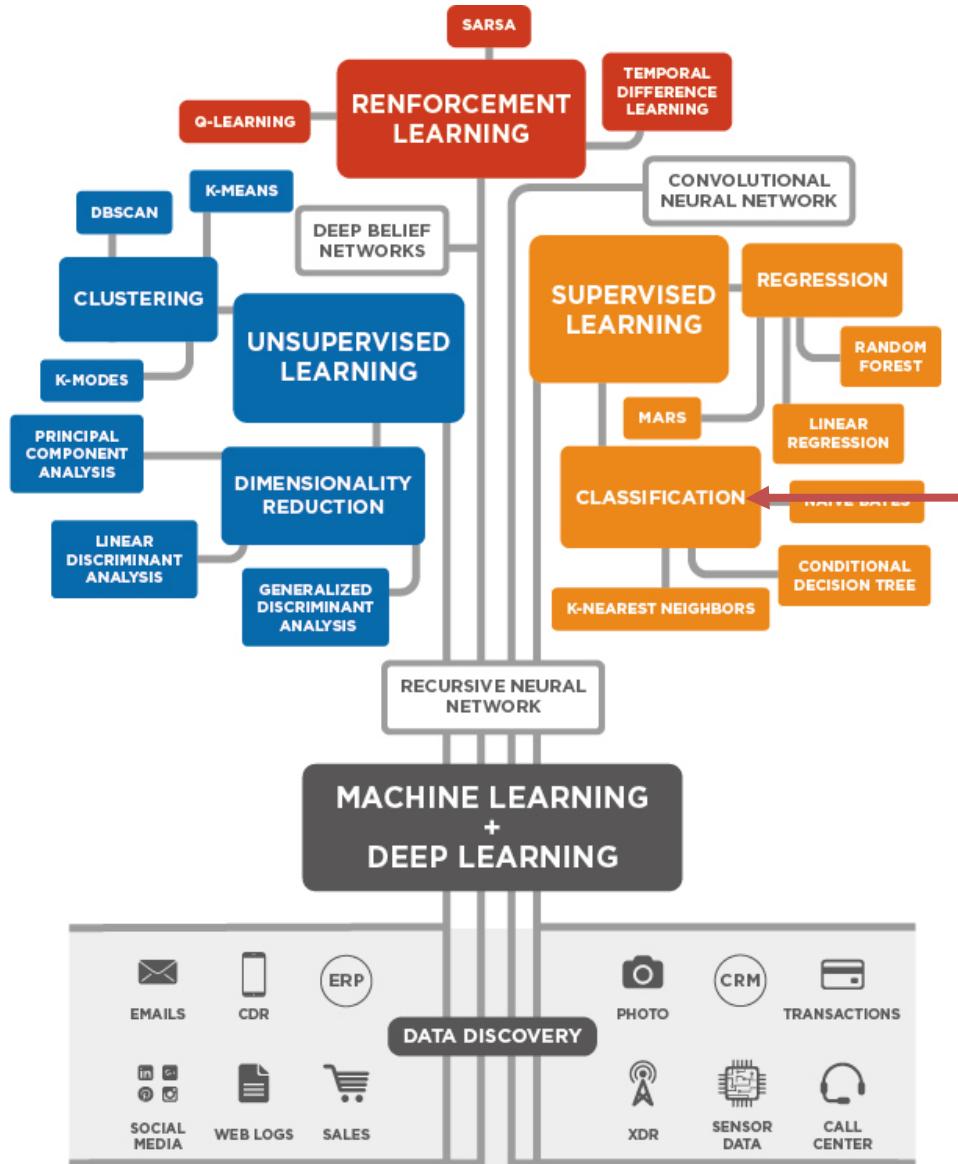
Anomaly Detection

Detection : Structure

Clustering and Dim-Red

Decision for Action

Reinforcement Learning



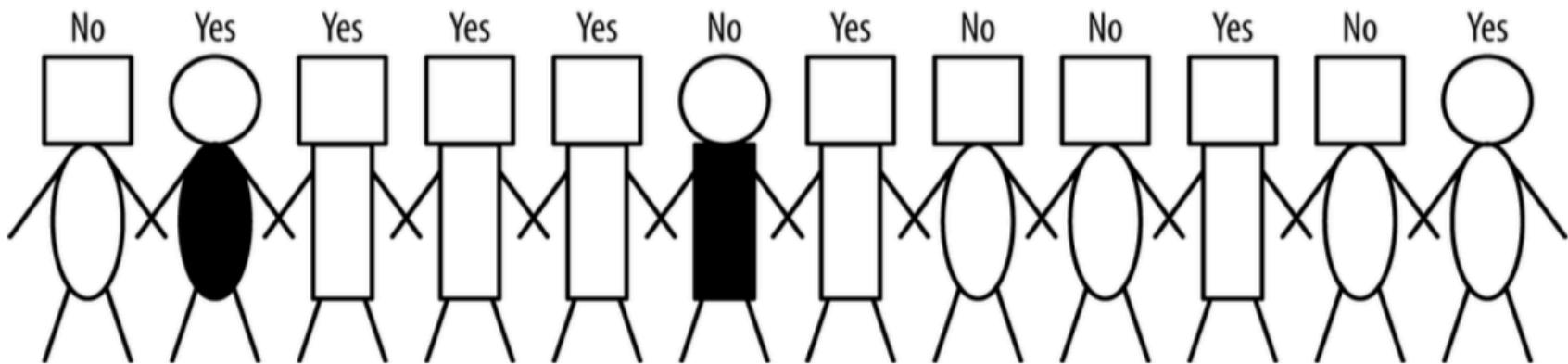
<http://blogs.teradata.com/data-points/tree-machine-learning-algorithms/>

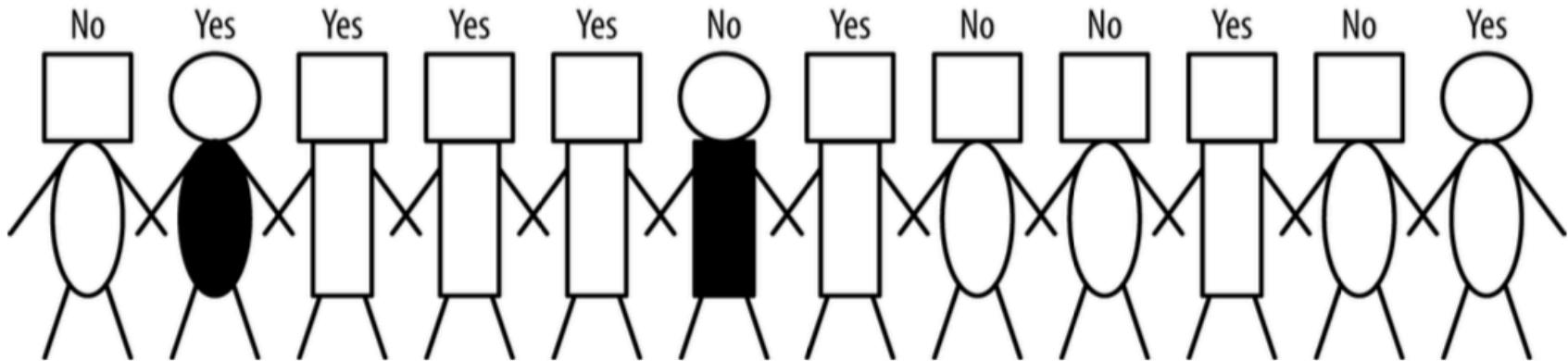


Is the credit rating “Good” or “Bad”?

HOW TO YOU DIFFERENTIATE?



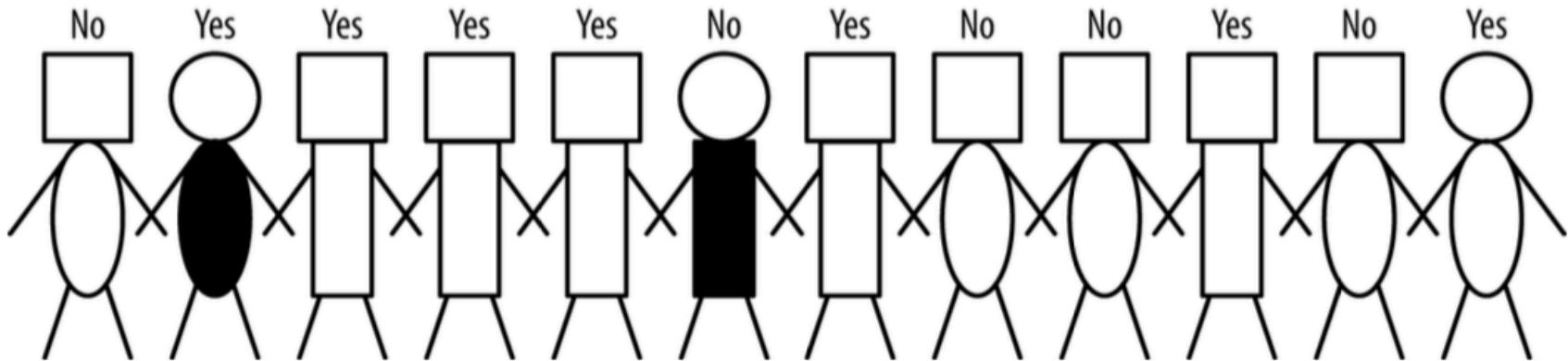




YES : NO

7 : 5

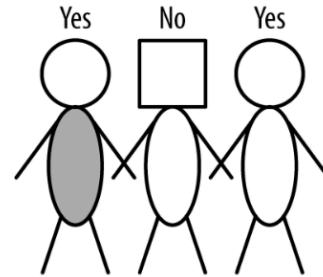
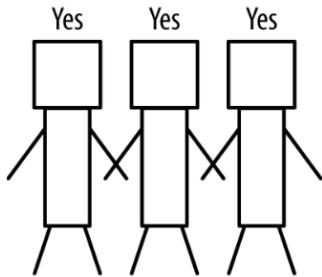




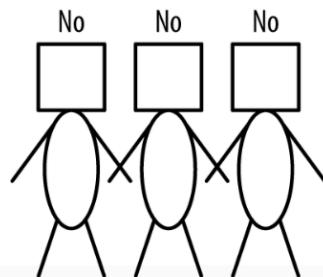
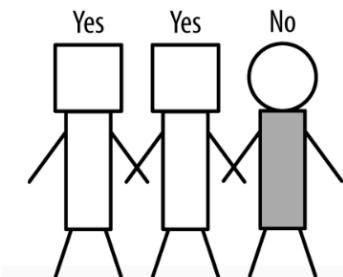
Rectangular Bodies

Oval Bodies

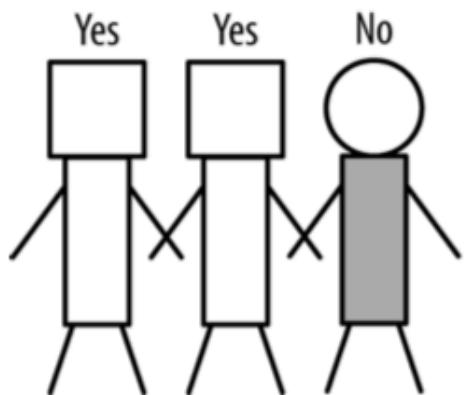
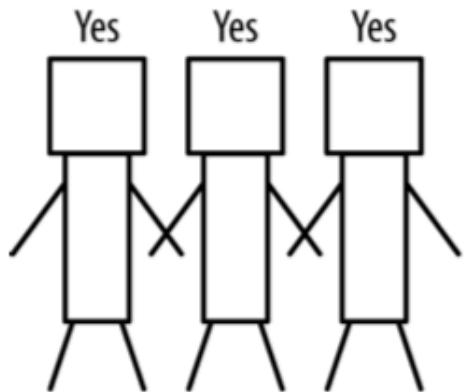
YES : NO
5 : 1



YES : NO
2 : 4



Rectangular Bodies

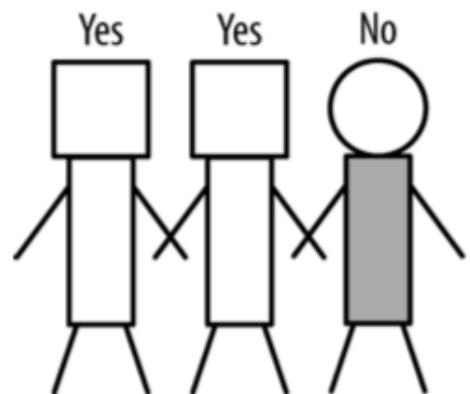
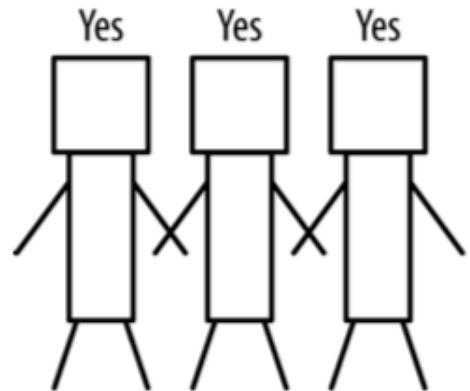


YES : NO

5 : 1



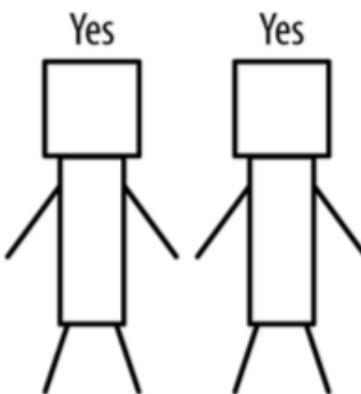
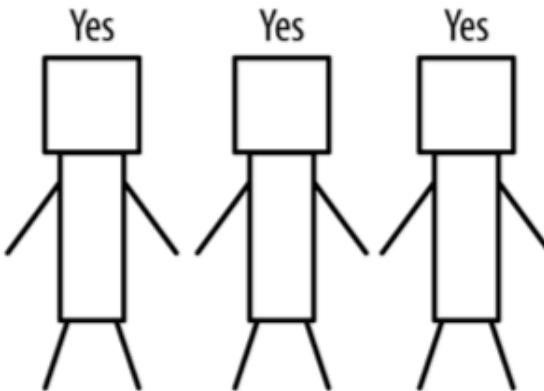
Rectangular Bodies



YES : NO

5 : 1

Rectangular Body and White



YES : NO

5 : 0

Rectangular Body and Gray

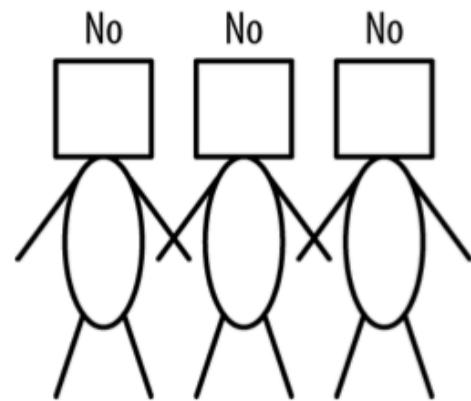
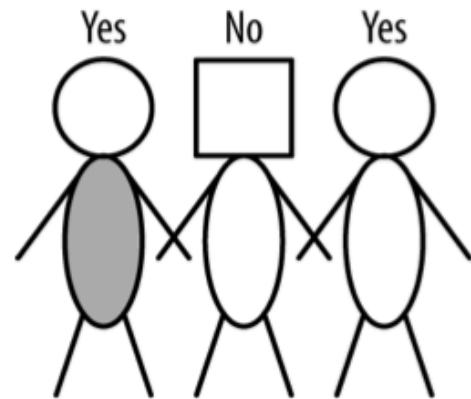


YES : NO

0 : 1



Oval Bodies

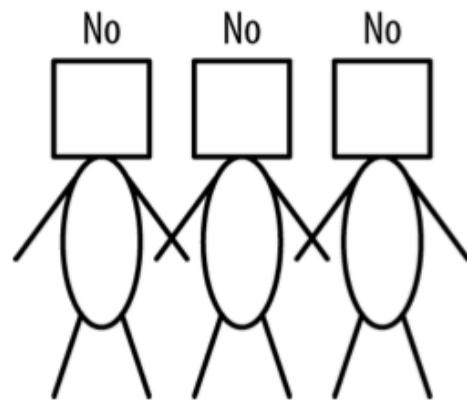
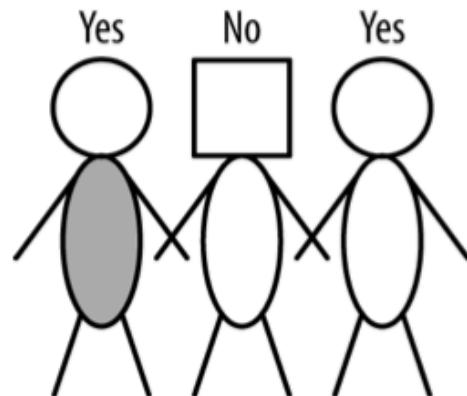


YES : NO

2 : 4



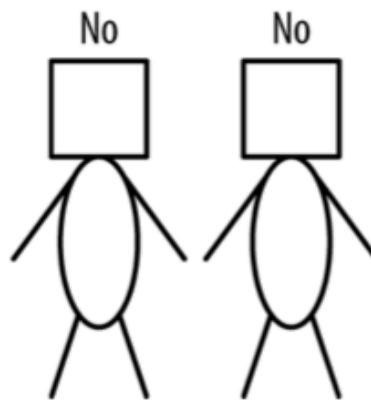
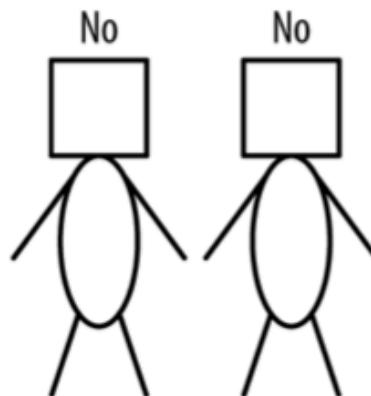
Oval Bodies



YES : NO

2 : 4

Oval Body and Square Head



YES : NO

0 : 4

Oval Body and Circular Head

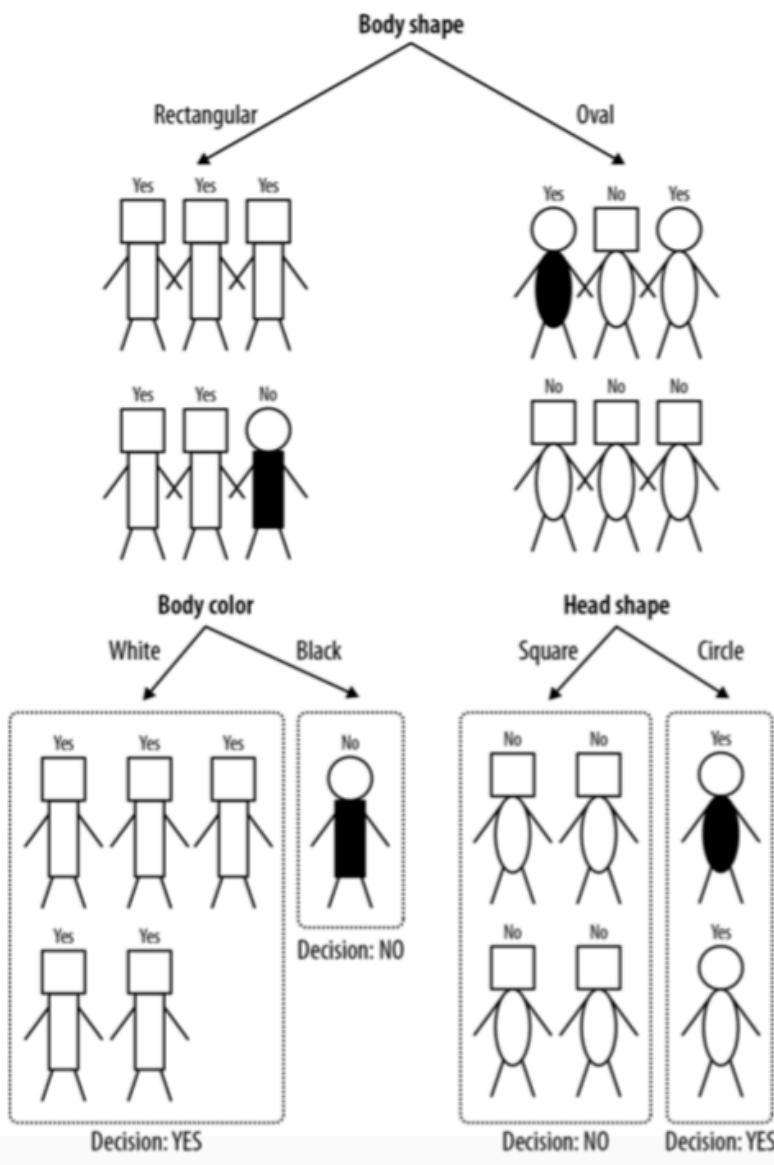


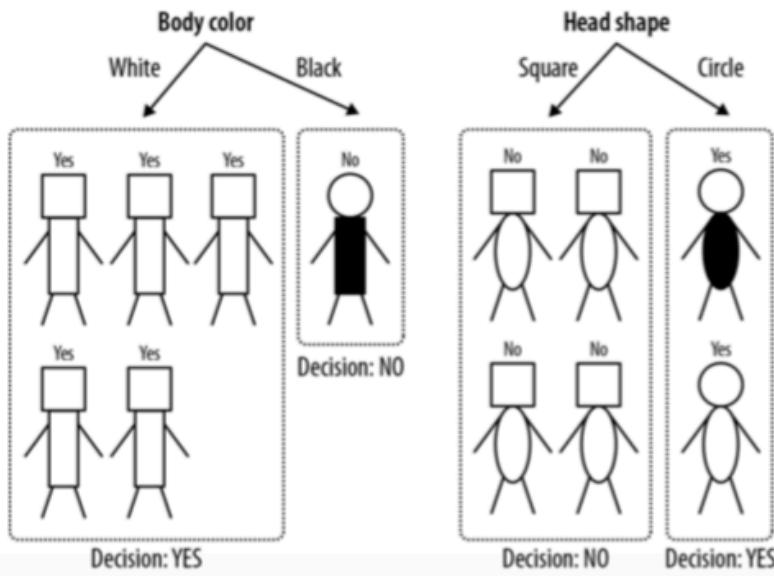
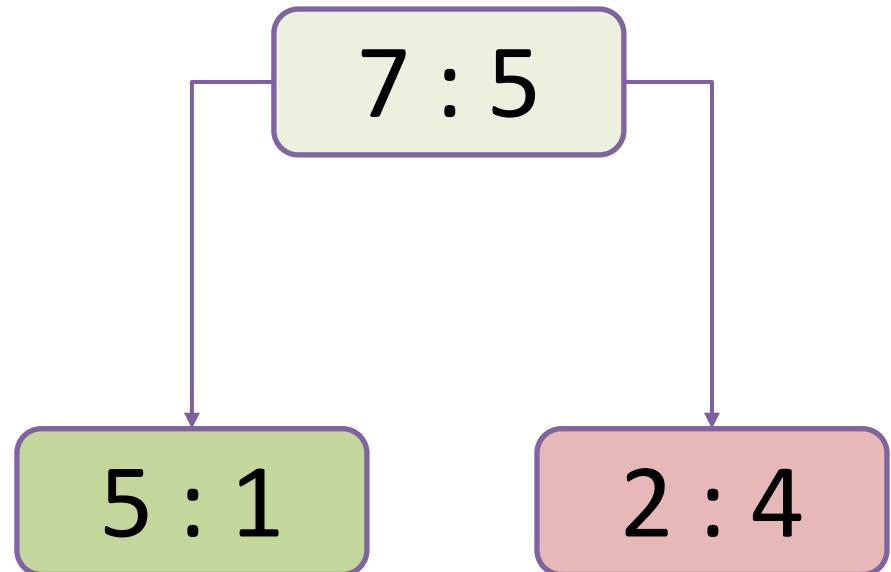
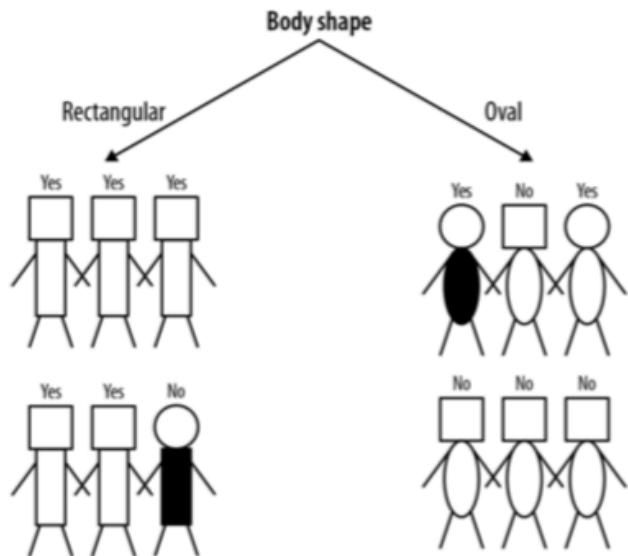
YES : NO

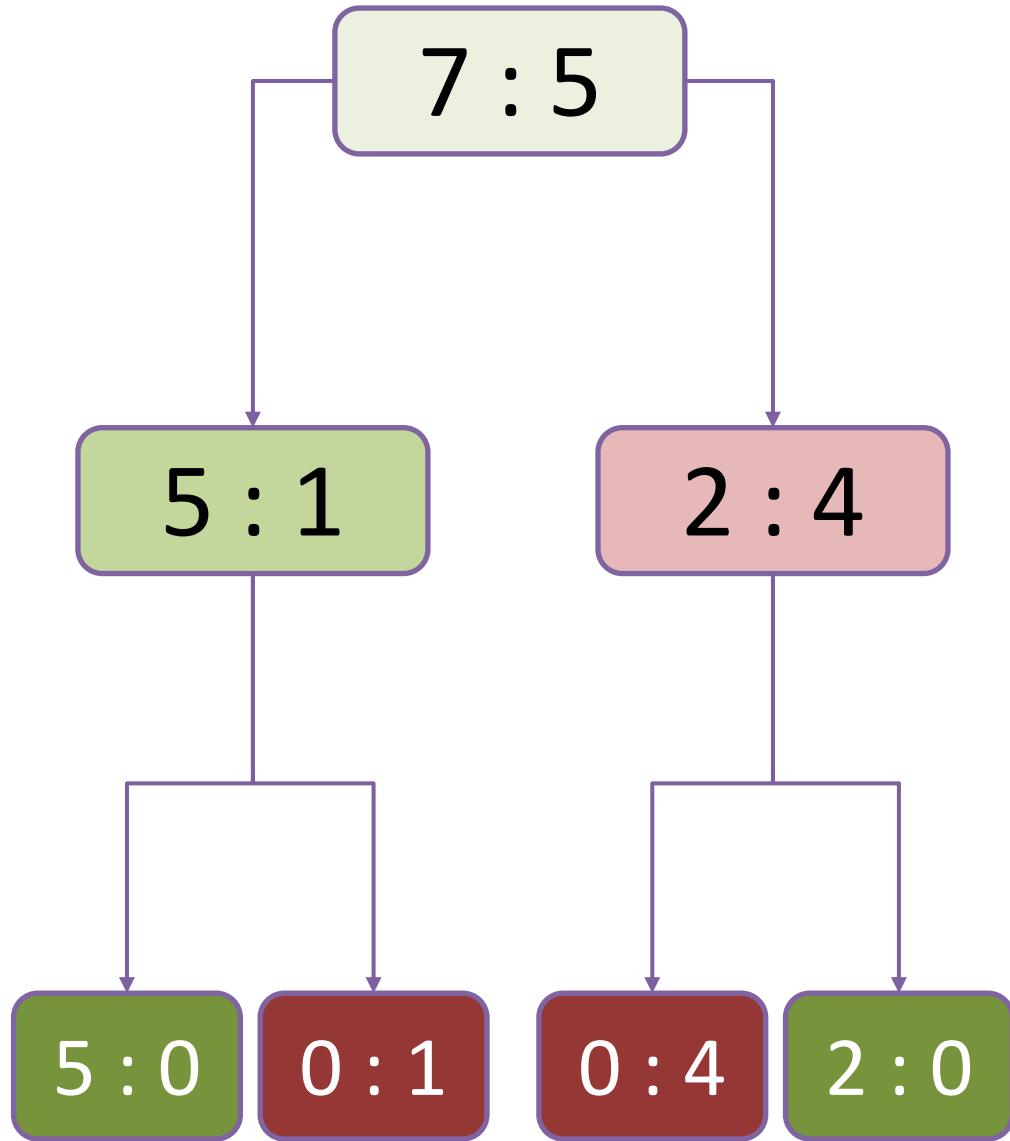
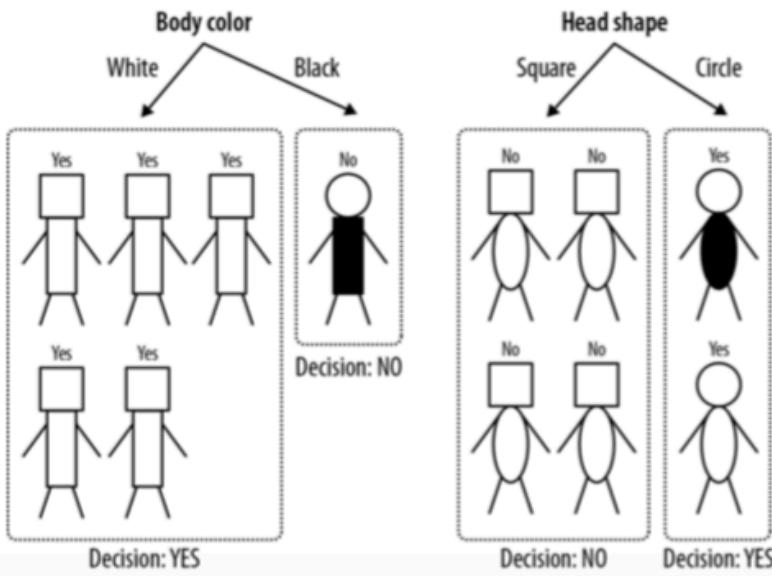
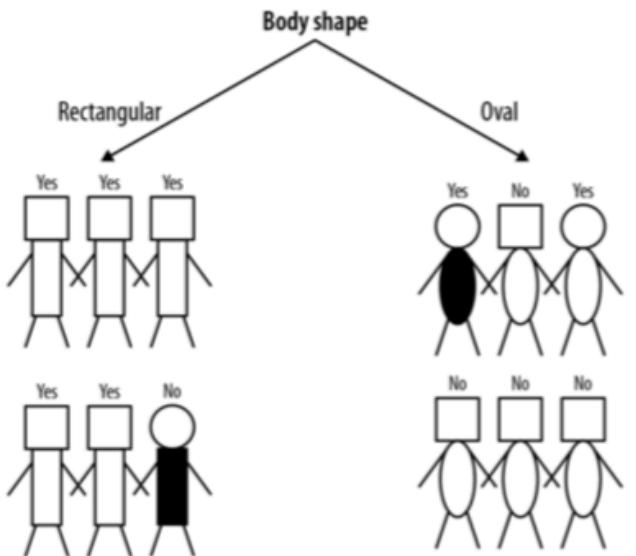
2 : 0

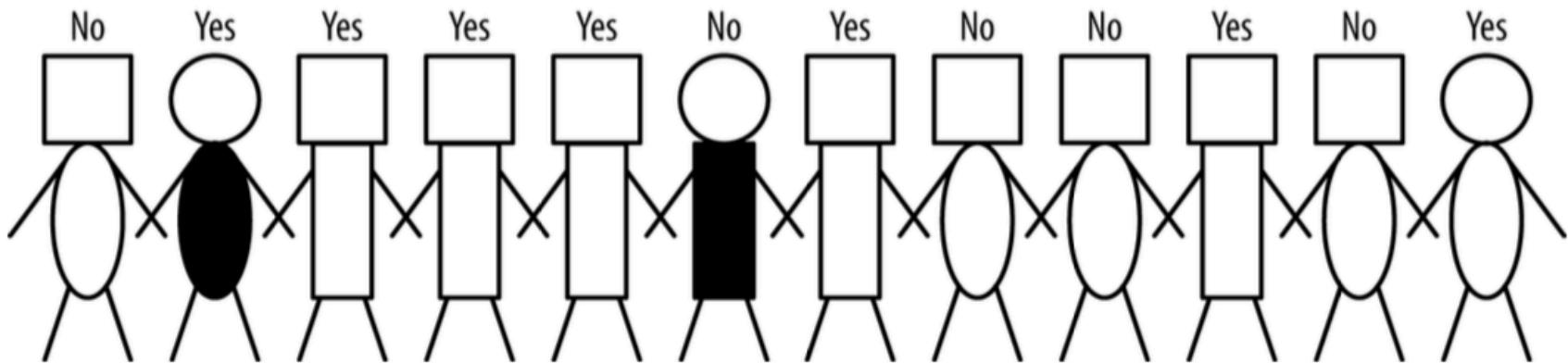


7 : 5





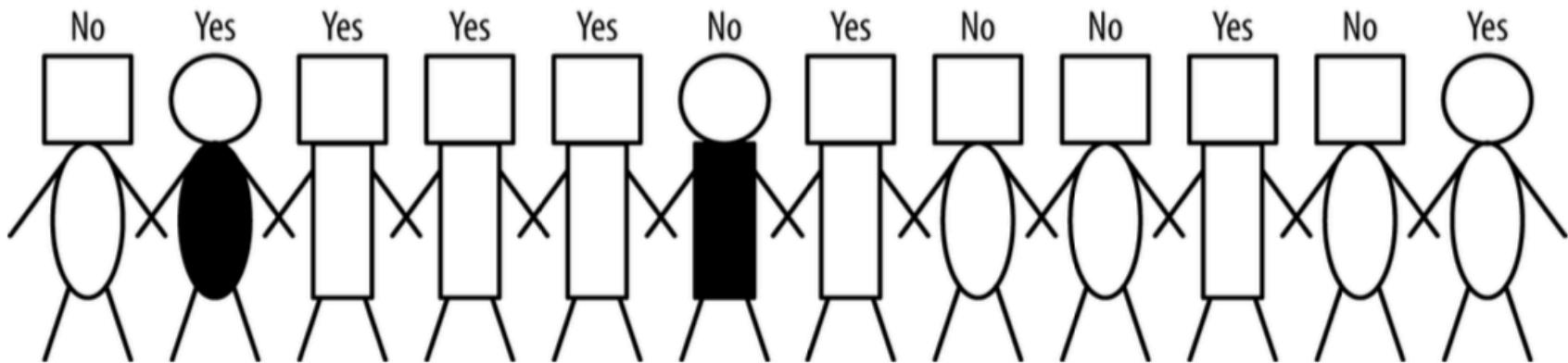




Body : Rectangle or Oval

Head : Square or Circle





Body : Rectangle or Oval

7 : 5

5 : 1

2 : 4

Head : Square or Circle

7 : 5

5 : 4

2 : 1



Prediction using a Decision Tree

HOW ACCURATE IS IT?



Confusion Matrix

	Predicted Bad	Predicted Good
Actually Bad	True Negatives	False Positives
Actually Good	False Negatives	True Positives



(0,1) Best

Neutral (1,1)

True Positive Rate

(0,0) Neutral

False Positive Rate

Worst (1,0)



(0,1) Best

Neutral (1,1)

True Positive Rate

The “No-Brainer” Model

(0,0) Neutral

False Positive Rate

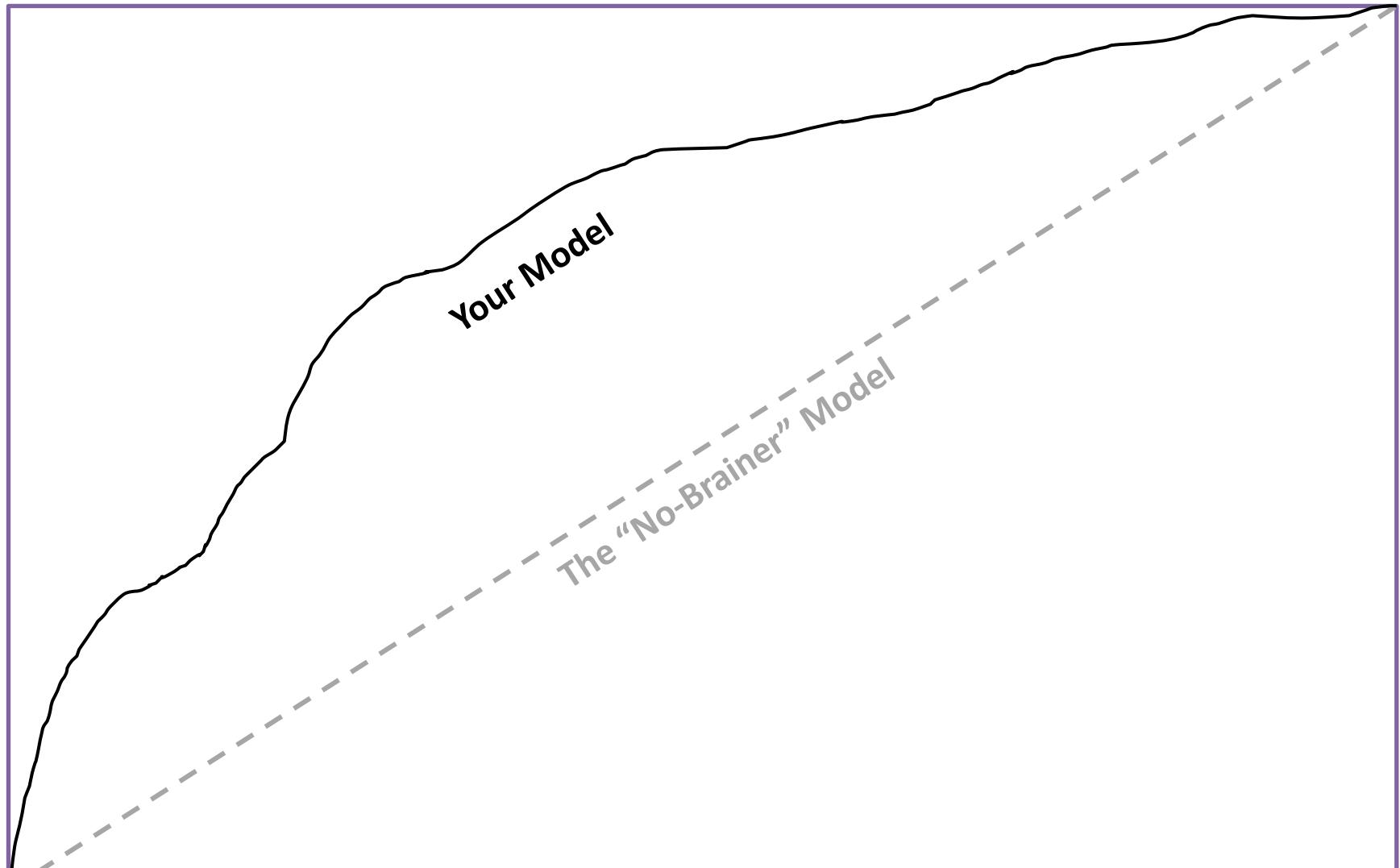
Worst (1,0)



(0,1) Best

Neutral (1,1)

True Positive Rate



(0,0) Neutral

False Positive Rate

Worst (1,0)

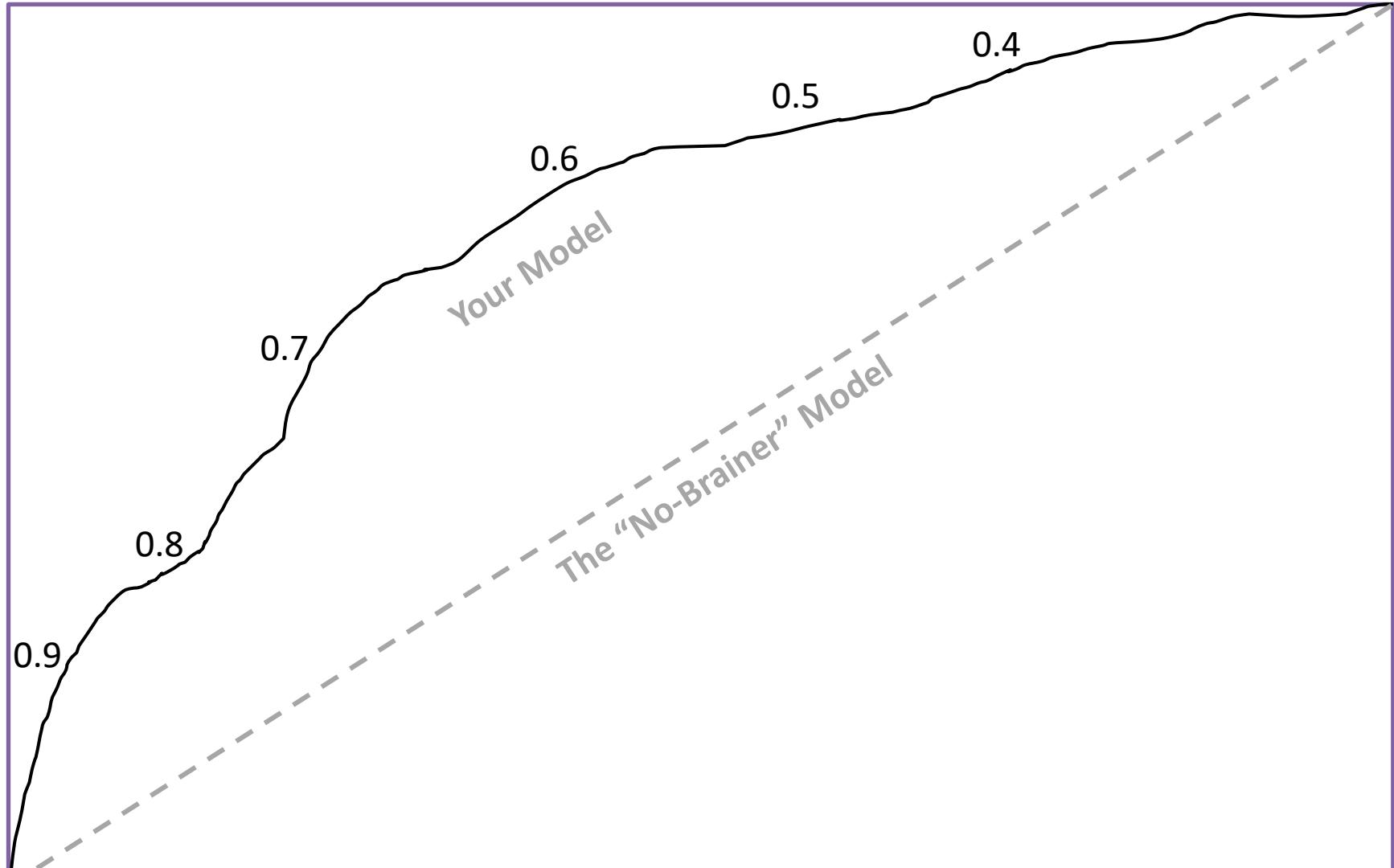


NANYANG TECHNOLOGICAL UNIVERSITY | SINGAPORE

(0,1) Best

Neutral (1,1)

True Positive Rate

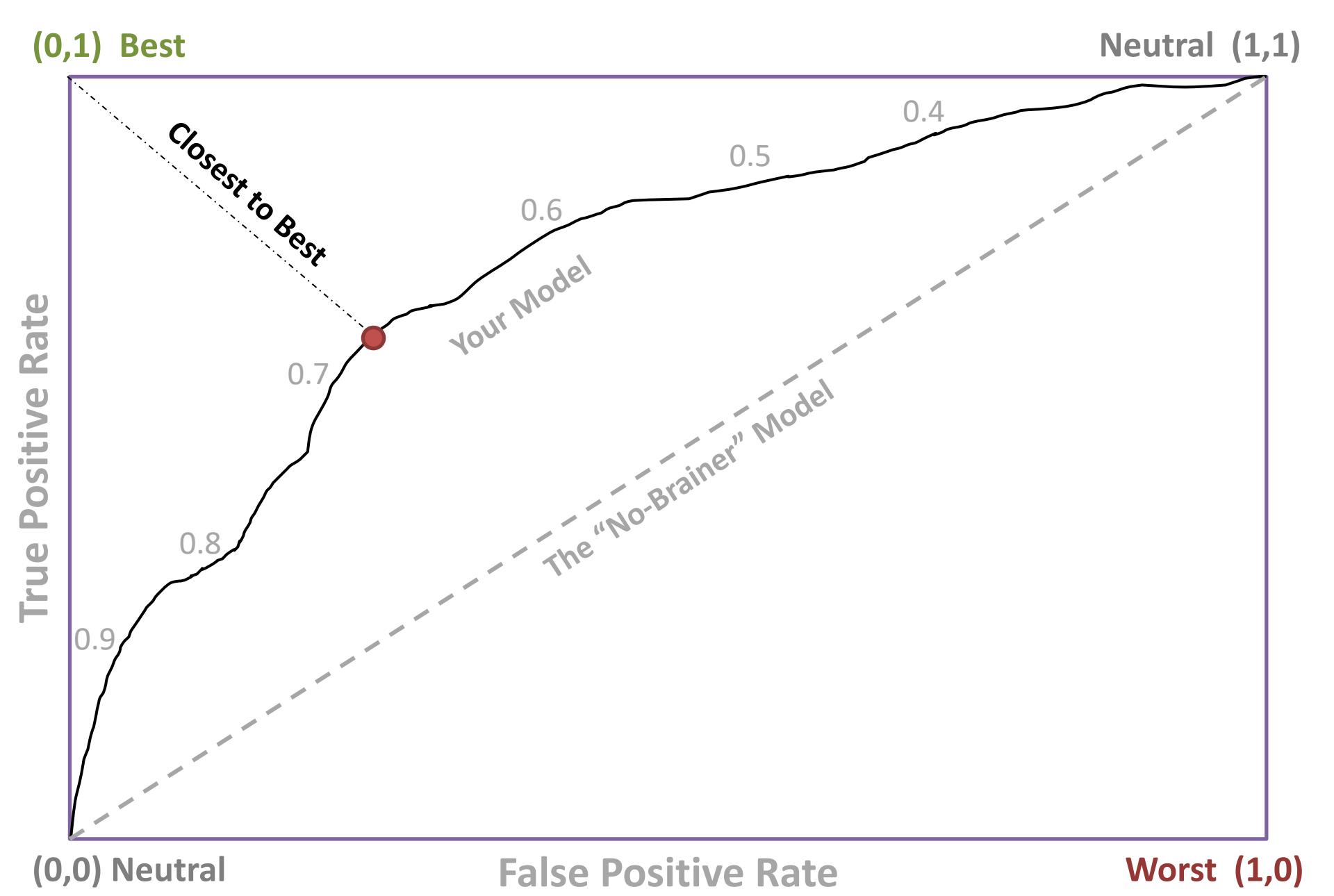


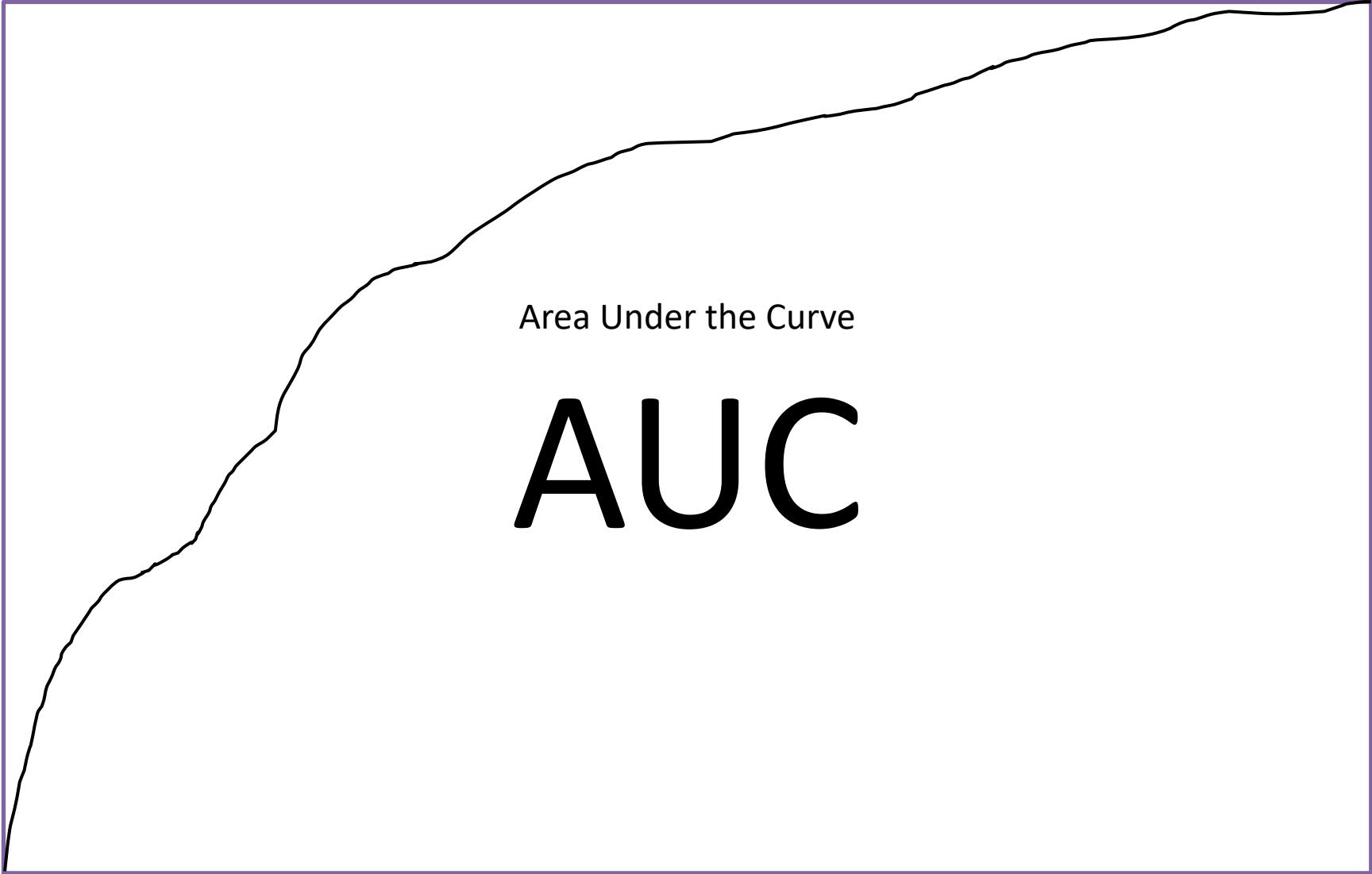
False Positive Rate



NANYANG TECHNOLOGICAL UNIVERSITY | SINGAPORE

Worst (1,0)





A graph showing the Area Under the Curve (AUC) for a classifier. The x-axis represents the False Positive Rate (FPR) and the y-axis represents the True Positive Rate (TPR). A black S-shaped curve starts at (0,0) and ends at (1,1), representing a perfect classifier. The area under this curve is shaded in light blue.

Area Under the Curve

AUC

