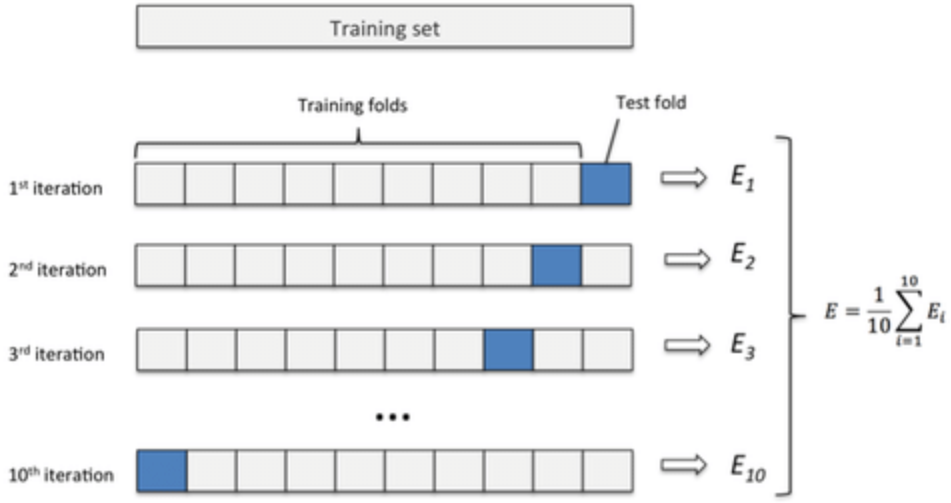
ROC Curves

n-Fold Cross Validation

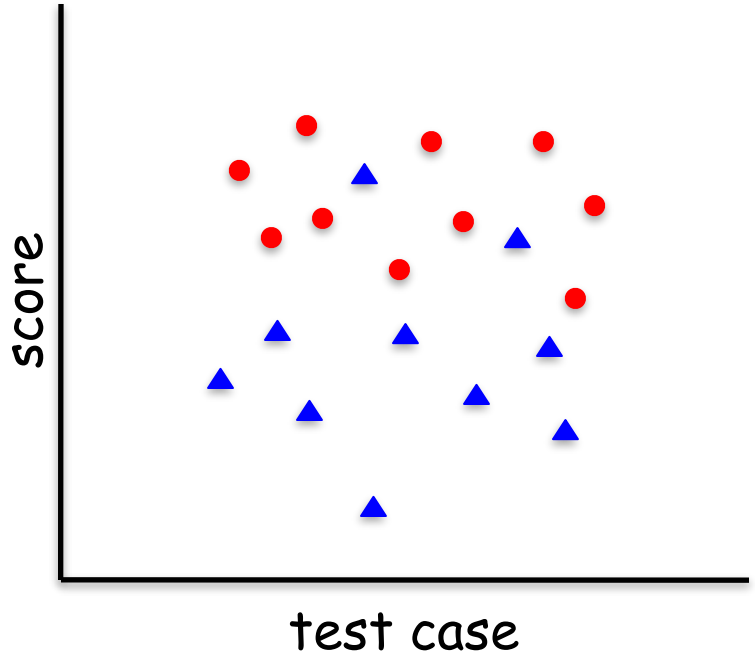
- Partition match set into n equal subsets
  - Denote subsets as $S_1, S_2, \ldots, S_n$
- Let training set be $S_2 \cup S_3 \cup \ldots \cup S_n$
  - And test set is S_1
- Repeat with training set $S_1 \cup S_3 \cup \ldots \cup S_n$
  - And test set S_2
- And so on, for each of n "folds"
  - Typically, n = 5 or n = 10 is used


Training set
Training folds
Test fold
1st iteration
2nd iteration
3rd iteration
10th iteration
E_1
E_2
E_3
E_{10}
$E = \frac{1}{10}\sum_{i=1}^{10} E_i$

Scatterplot

- Train a model on the training set
- Apply score to test
  - Can visualize results as a scatterplot

- match scores
- **nomatch** scores

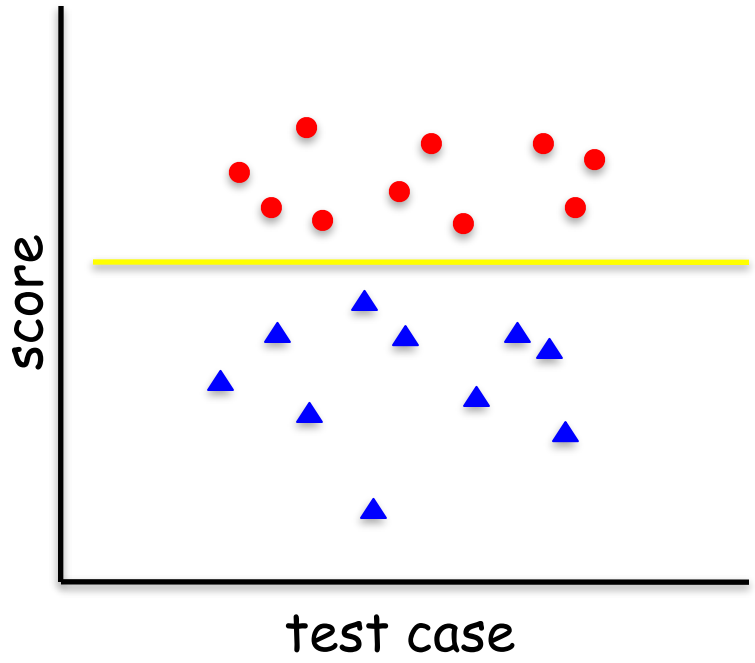


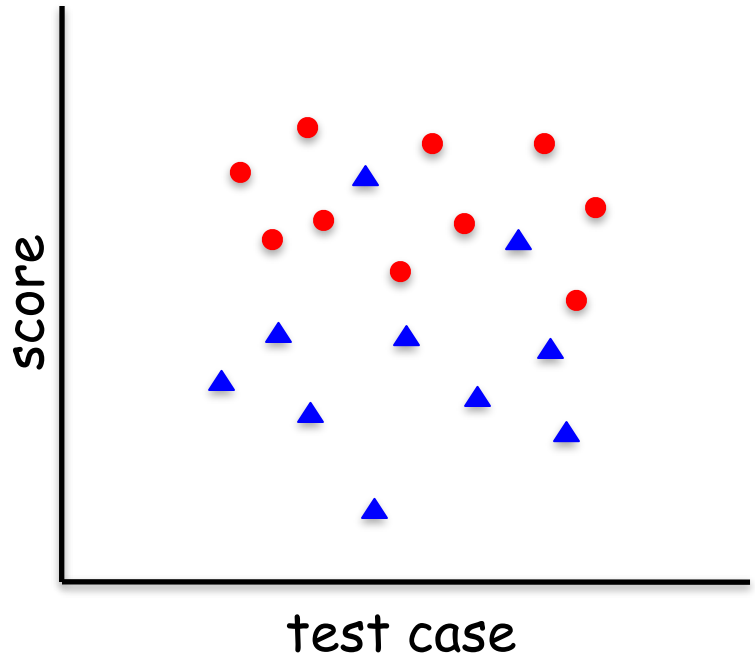
Thresholding

- Set threshold after scoring phase
- Ideally, we have complete separation
  - I.e., no "overlap" in scatterplot
  - Usually, that doesn't happen
  - So, where to set the threshold?
- In practice, thresholding is critical

Thresholding

- Where to set threshold?
  - Left scatterplot is a lot easier than right



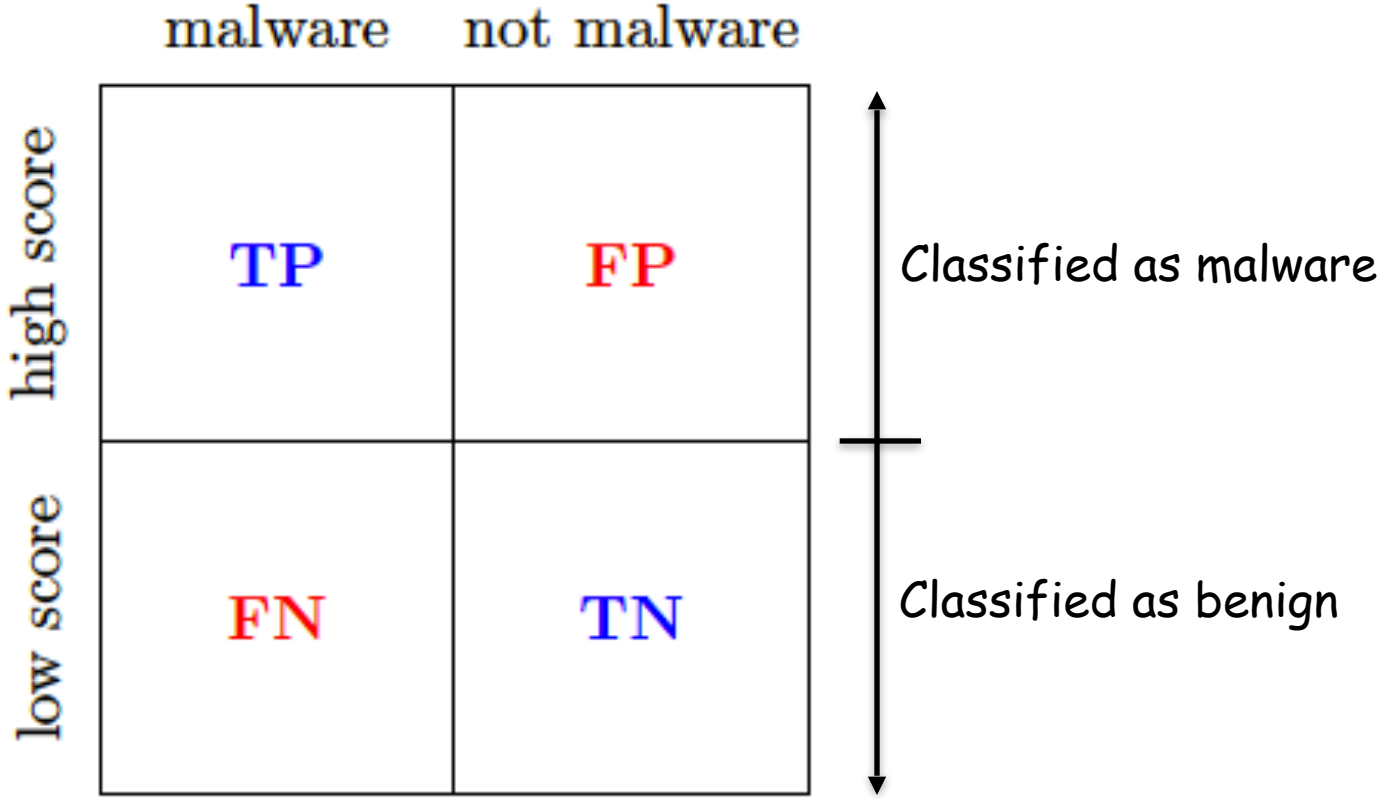


Results

- Given scatterplot and a threshold
- For each sample, one of 4 cases…
  - **True positive** — correctly classified as +
  - **False positive** — incorrectly classified as +
  - **True negative** — correctly classified as –
  - **False negative** — incorrectly classified as –
- TP, FP, TN, FN, respectively
  - Append "R" to each for "rate"

Confusion Matrix

- Assuming that high scores (i.e., above threshold) better match the model



Sensitivity and Specificity

- The **TPR** also known as **sensitivity** while **TNR** is known as **specificity**
- Consider a medical test
  - Sensitivity is percentage of sick people detected by the test (as they should be)
  - Specificity is percentage of healthy who are not classified as sick (as they should)
- Inherent tradeoff between TPR & TNR
  - Everything depends on threshold!

Accuracy

- Let P be the number of positive cases tested and N the negative cases tested
  - Note: P is size of test set, N nomatch set
  - Also, **P = TP + FN** and **N = TN + FP**
- Then, **Accuracy = (TP + TN) / (P + N)**
  - Note that accuracy ranges from 0 to 1
  - Accuracy of 1? Ideal situation
  - Accuracy 0.5? Don't give up your day job…

Accuracy

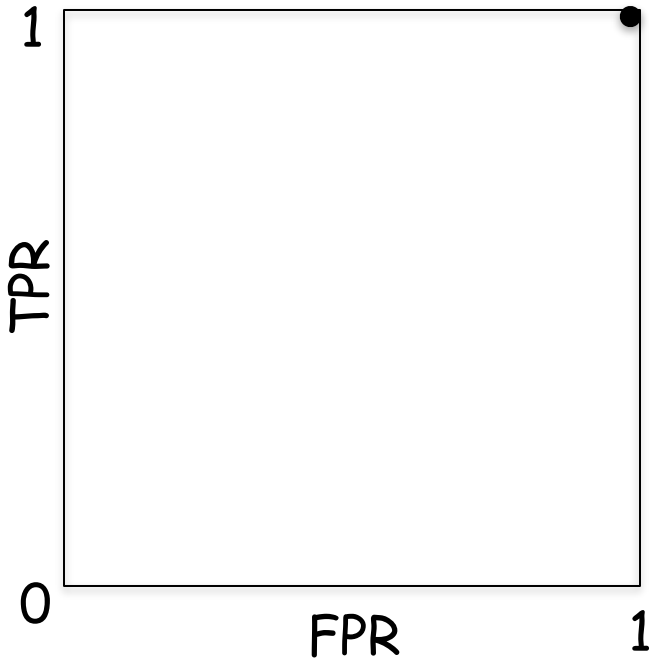
- Accuracy tells us something…
  - But it **depends on where threshold is set**
  - How should we set the threshold?
  - Seems we are going around in circles, like a dog chasing its tail
- Bottom line? We still don't have good way to compare different techniques
  - Next slide, please…

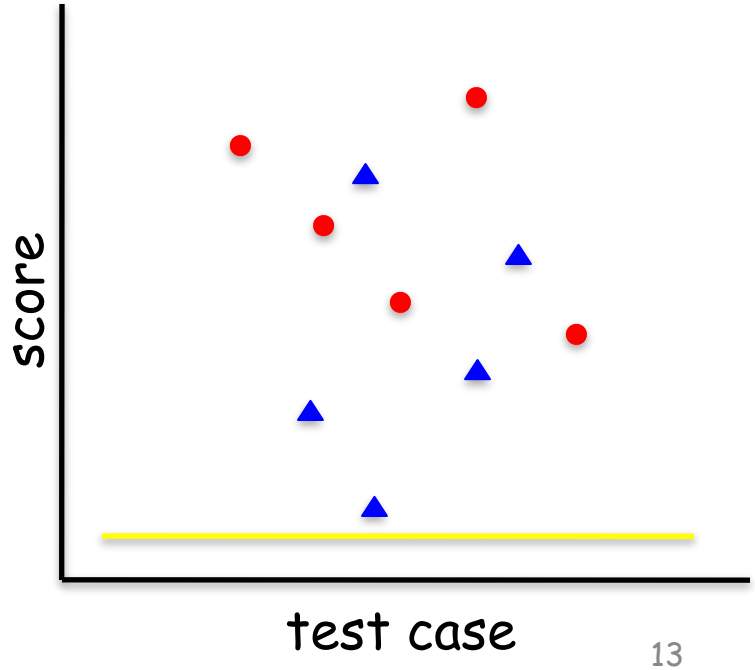
ROC Curves

- **Receiver Operating Characteristic**
  - Originated from electrical engineering
  - Now widely used in many fields
- What is an **ROC curve**?
  - ➢Plot TPR vs FPR as threshold varies thru the range of scores
  - Plot FPR on x-axis, TPR on y-axis
  - Equivalently, 1 – specificity vs sensitivity
- What the … ?

ROC Curve

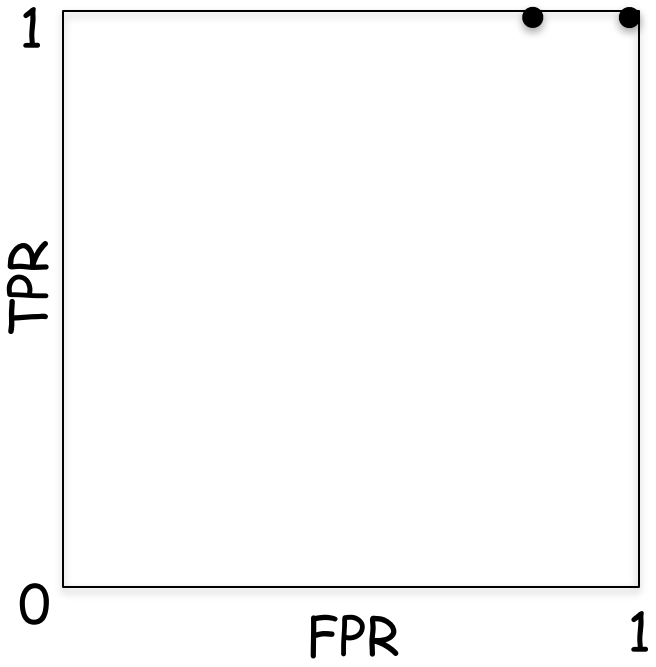
- Suppose threshold set at yellow line
  - Above yellow, classified as positive,
  - Below yellow is negative
- In this case,
  - TPR = 1.0
  - FPR = 1.0 – TNR = 1.0 – 0.0 = 1.0

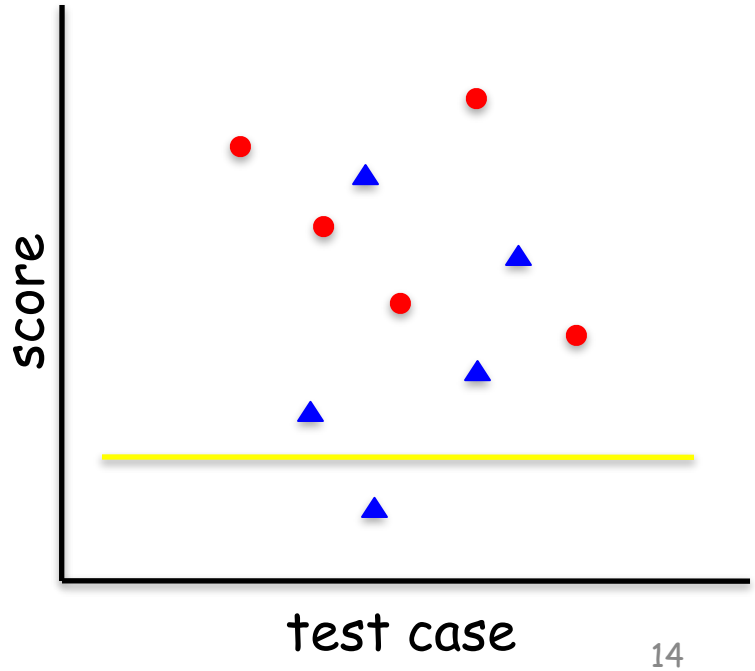




ROC Curve

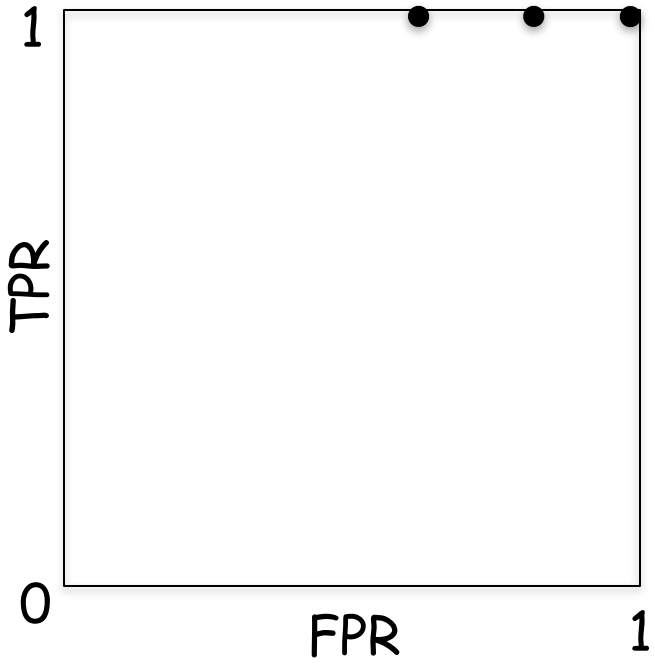
- Suppose threshold set at yellow line
  - Above yellow, classified as positive,
  - Below yellow is negative
- In this case,
  - TPR = 1.0
  - FPR = 1.0 – TNR = 1.0 – 0.2 = 0.8

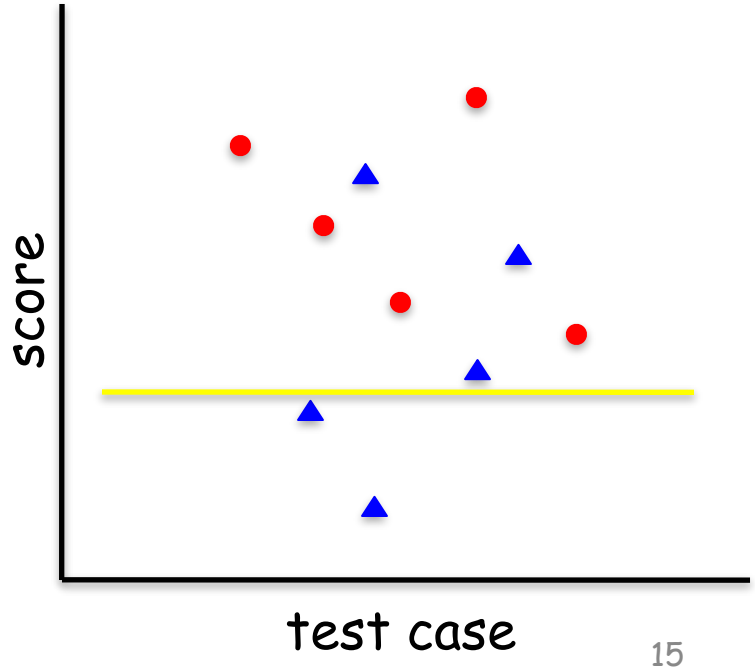




ROC Curve

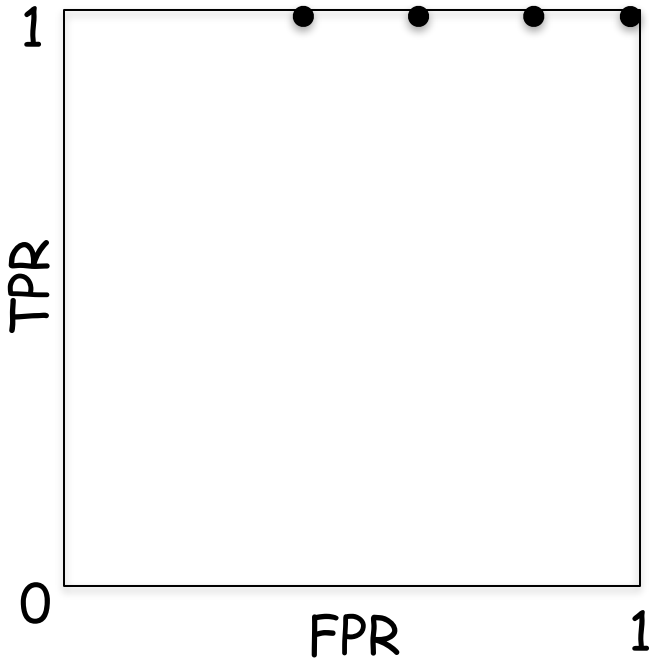
- Suppose threshold set at yellow line
  - Above yellow, classified as positive,
  - Below yellow is negative
- In this case,
  - TPR = 1.0
  - FPR = 1.0 – TNR = 1.0 – 0.4 = 0.6

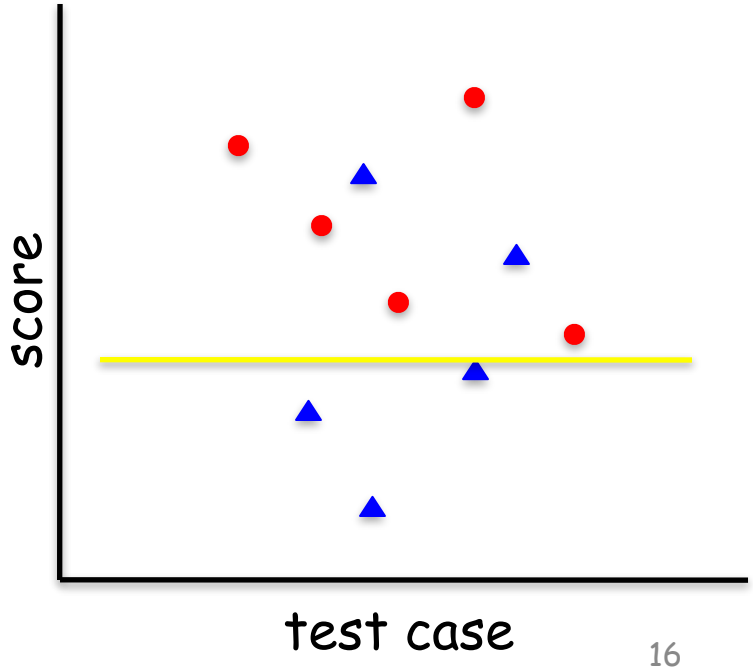




ROC Curve

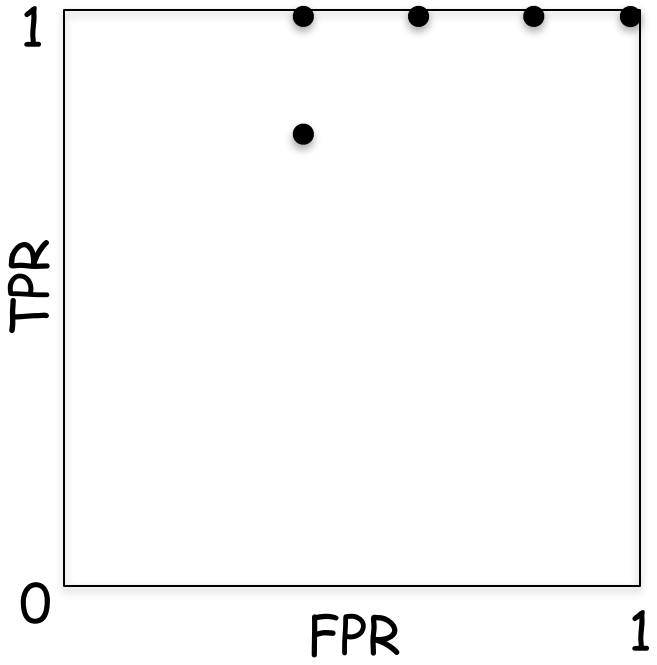
- Suppose threshold set at yellow line
  - Above yellow, classified as positive,
  - Below yellow is negative
- In this case,
  - TPR = 1.0
  - FPR = 1.0 – TNR = 1.0 – 0.6 = 0.4

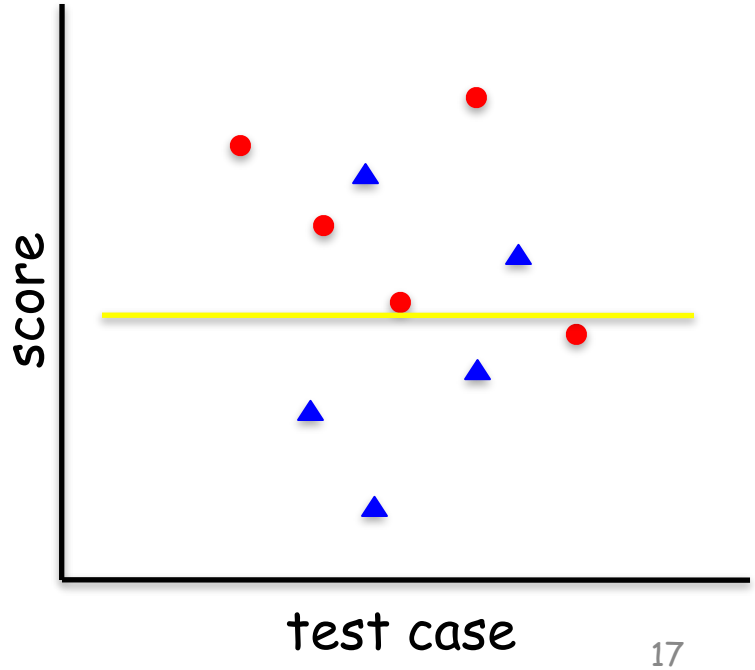




ROC Curve

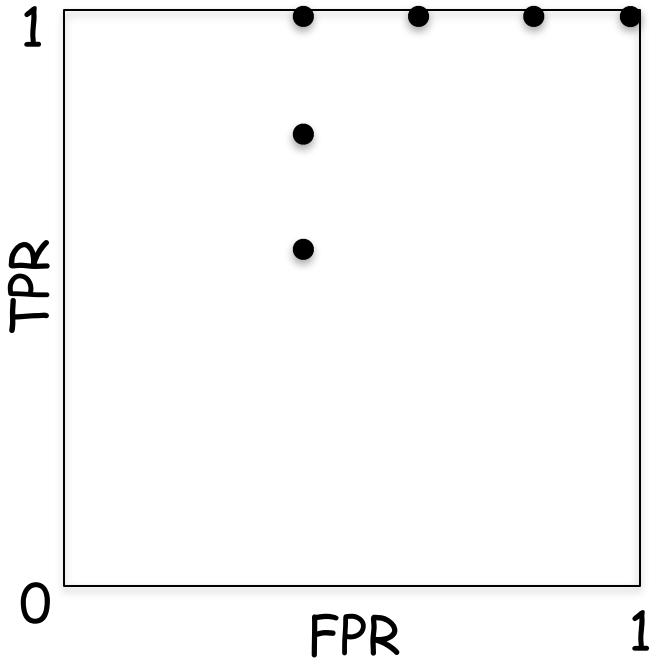
- Suppose threshold set at yellow line
  - Above yellow, classified as positive
  - Below yellow is negative
- In this case,
  - TPR = 0.8
  - FPR = 1.0 – TNR = 1.0 – 0.6 = 0.4

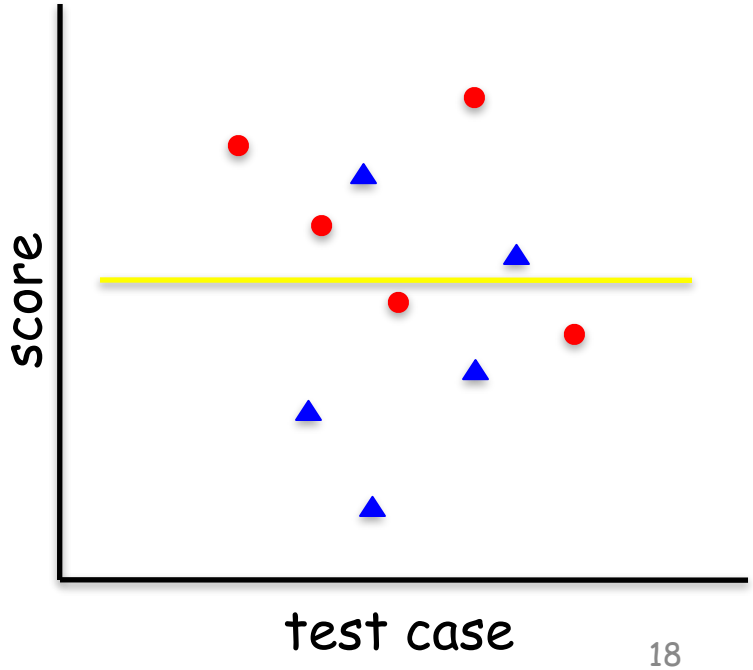




ROC Curve

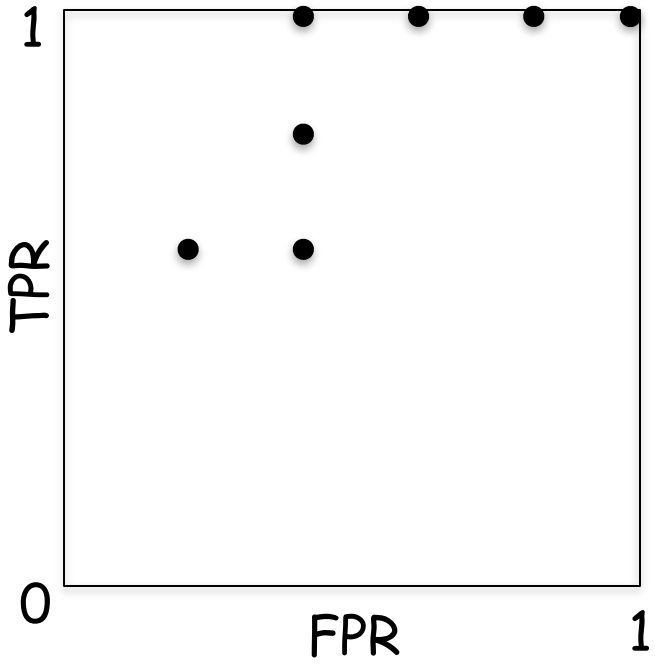
- Suppose threshold set at yellow line
  - Above yellow, classified as positive,
  - Below yellow is negative
- In this case,
  - TPR = 0.6
  - FPR = 1.0 – TNR = 1.0 – 0.6 = 0.4

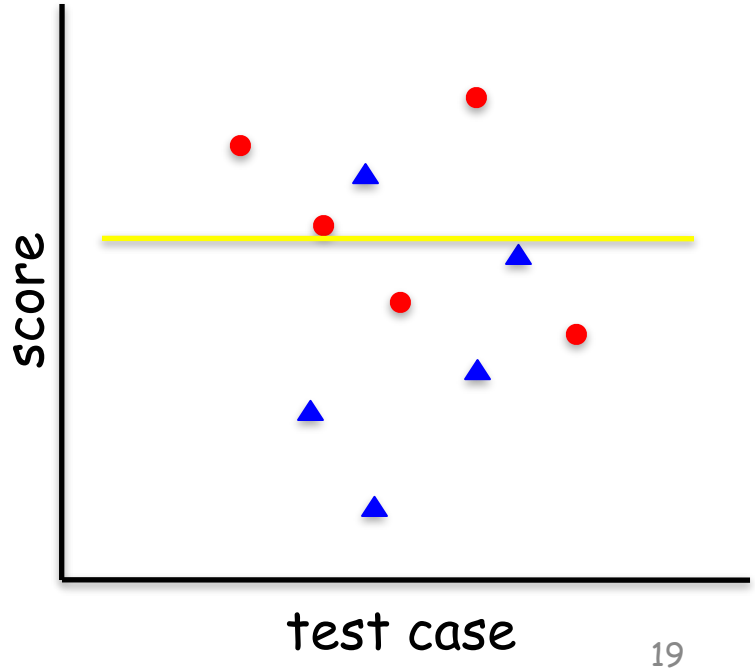




ROC Curve

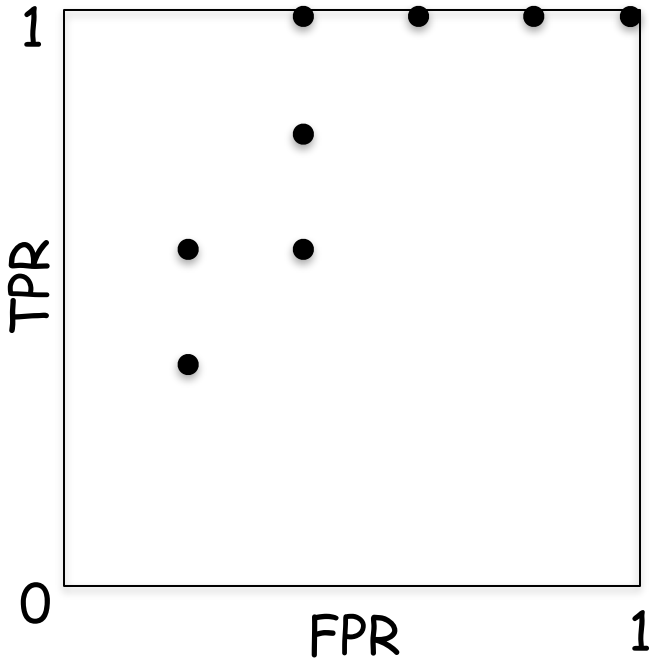
- Suppose threshold set at yellow line
  - Above yellow, classified as positive,
  - Below yellow is negative
- In this case,
  - TPR = 0.6
  - FPR = 1.0 – TNR = 1.0 – 0.6 = 0.2

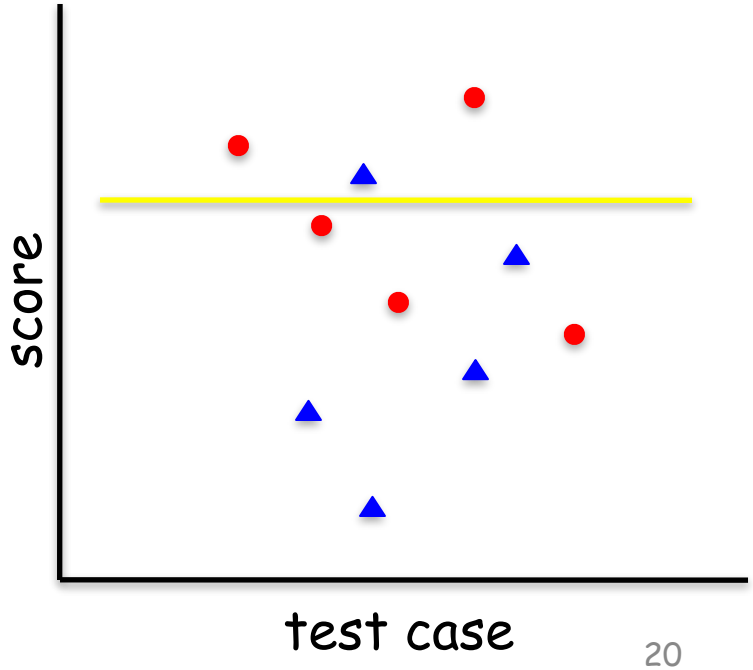




ROC Curve

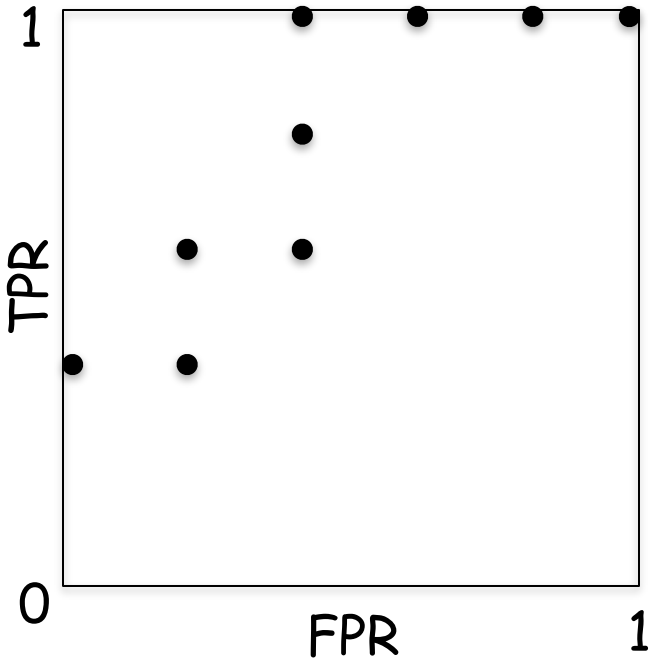
- Suppose threshold set at yellow line
  - Above yellow, classified as positive,
  - Below yellow is negative
- In this case,
  - TPR = 0.4
  - FPR = 1.0 – TNR = 1.0 – 0.6 = 0.2

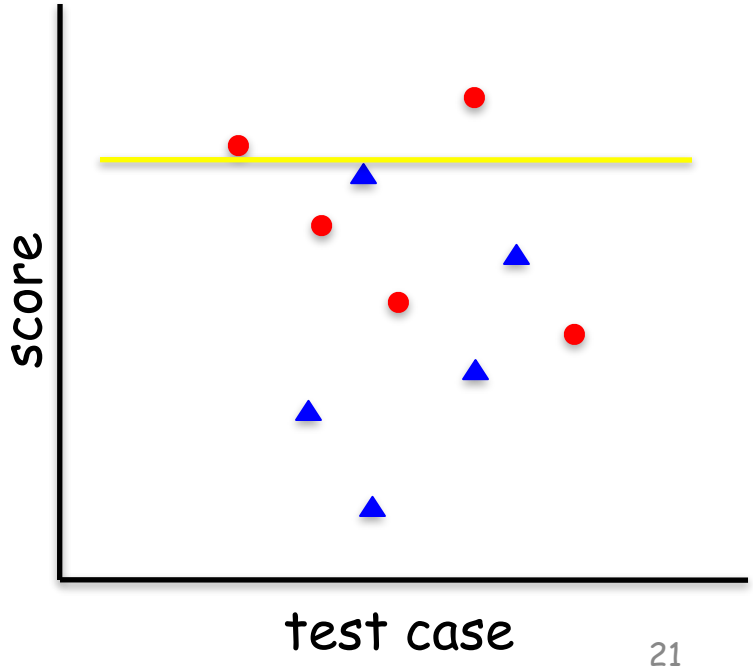




ROC Curve

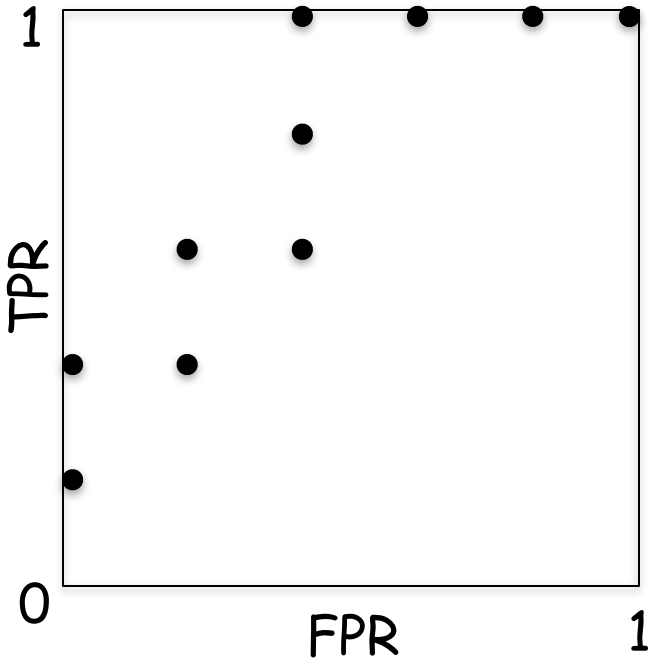
- Suppose threshold set at yellow line
  - Above yellow, classified as positive,
  - Below yellow is negative
- In this case,
  - TPR = 0.4
  - FPR = 1.0 – TNR = 1.0 – 1.0 = 0.0

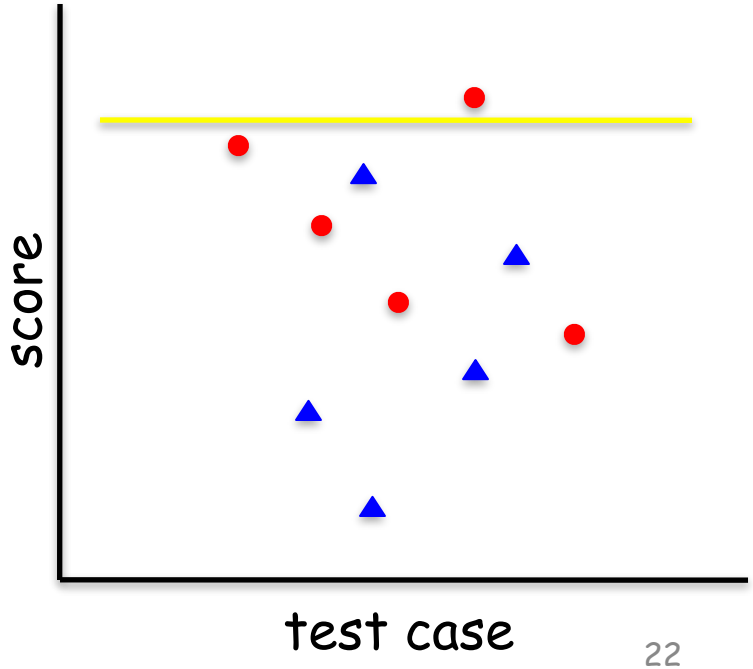




ROC Curve

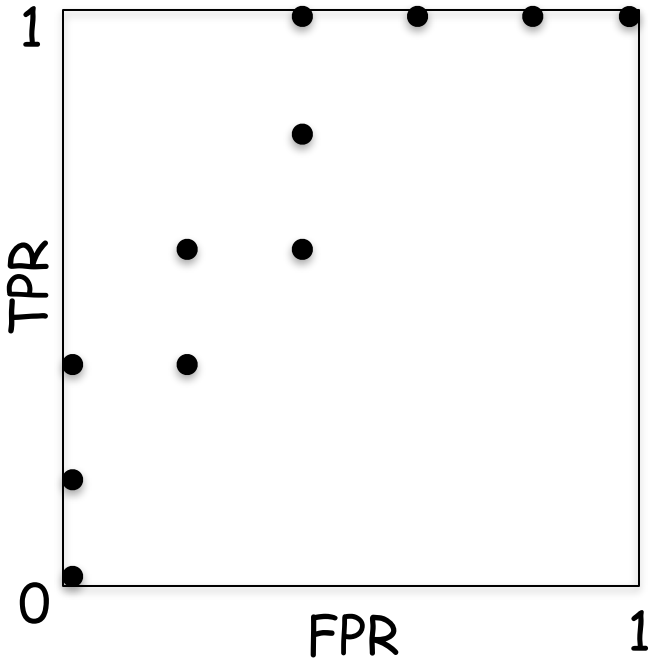
- Suppose threshold set at yellow line
  - Above yellow, classified as positive,
  - Below yellow is negative
- In this case,
  - TPR = 0.2
  - FPR = 1.0 – TNR = 1.0 – 0.6 = 0.0

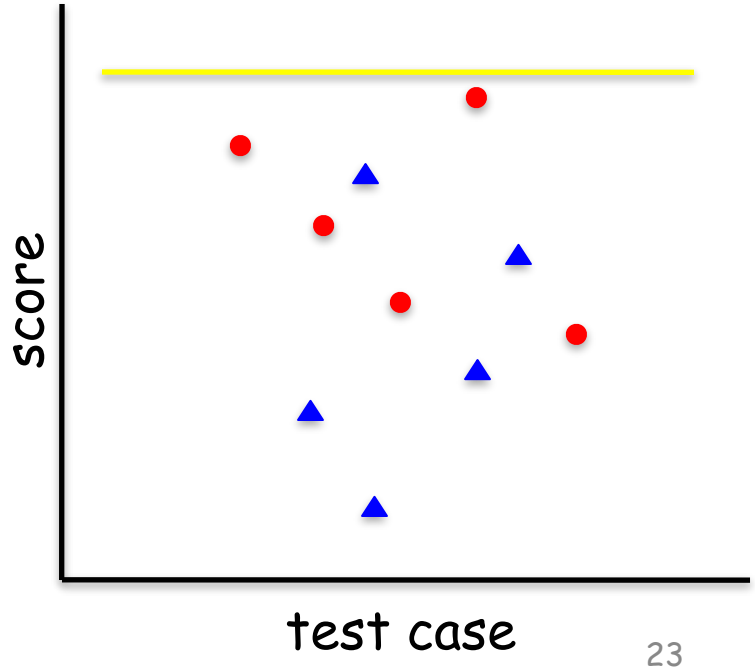




ROC Curve

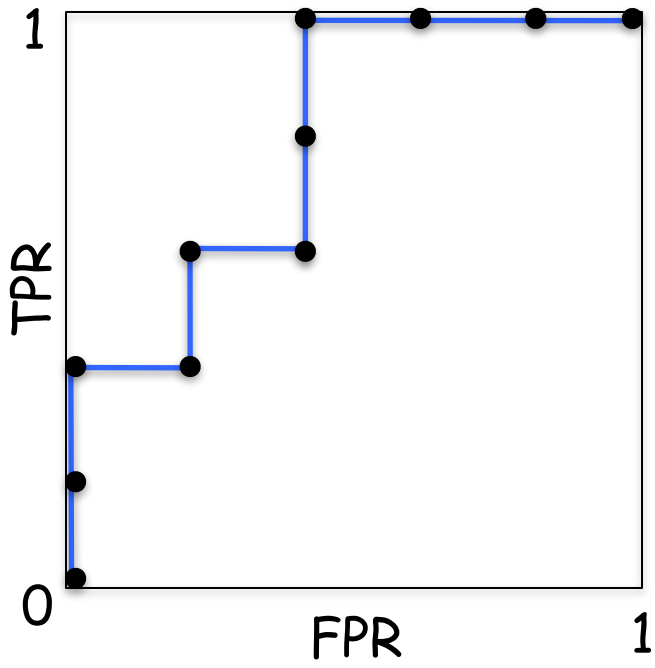
- Suppose threshold set at yellow line
  - Above yellow, classified as positive,
  - Below yellow is negative
- In this case,
  - TPR = 0.0
  - FPR = 1.0 – TNR = 1.0 – 0.6 = 0.0





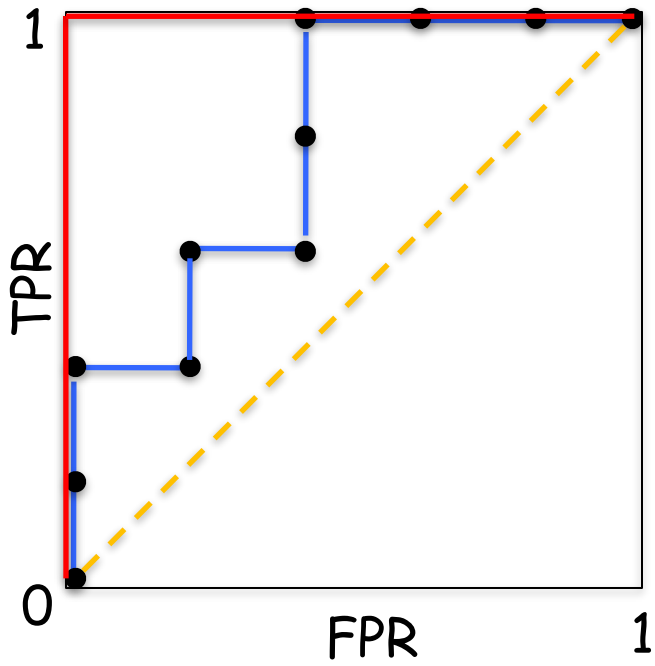
ROC Curve

- Connect the dots…
- This is a ROC curve!
- What good is it?
  - Captures info wrt **all possible** thresholds
  - Removes threshold as a factor in the analysis
- What does it all mean?



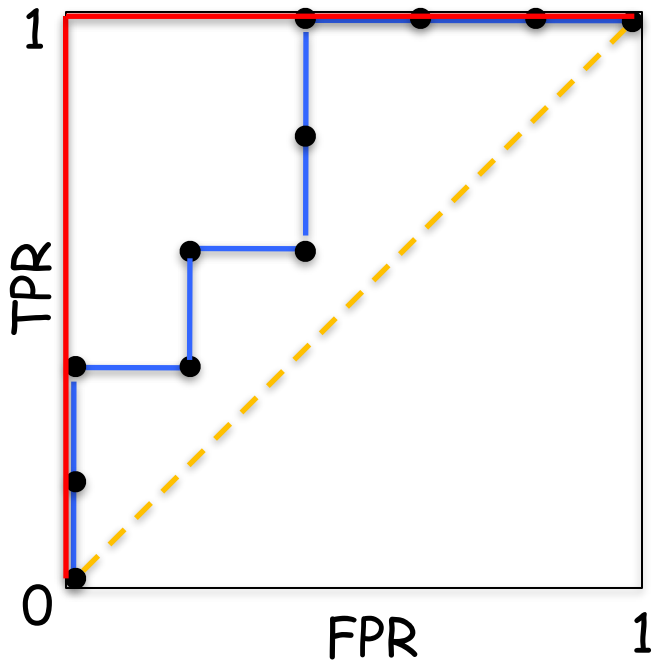
ROC Curve

- Random classifier?
  - Orange 45 degree line
- Perfect classifier?
  - Red (Why?)
- Above 45 degree line?
  - Better than random
  - The closer to the red, the closer to ideal



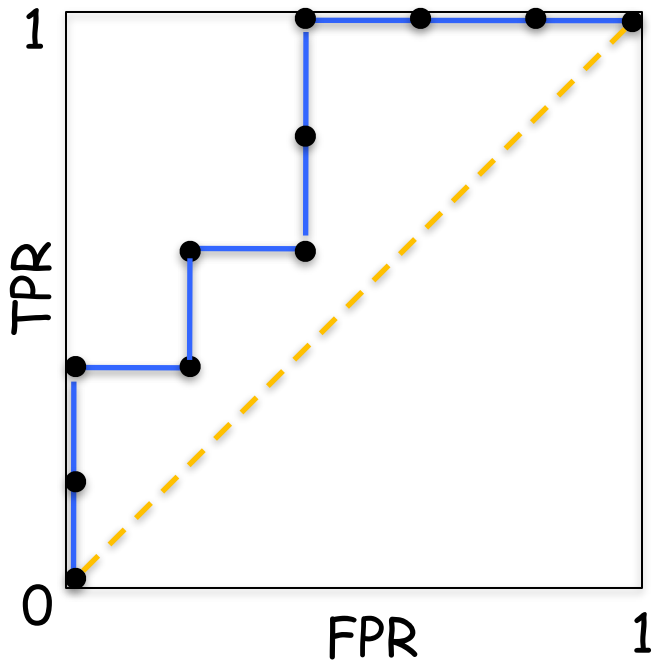
Area Under the Curve (AUC)

- ROC curve lives within a 1x1 square
- Random classifier?
  - AUC ≈ 0.5
- Perfect classifier (red)?
  - AUC = 1.0
- Example curve (blue)?
  - AUC = 0.8



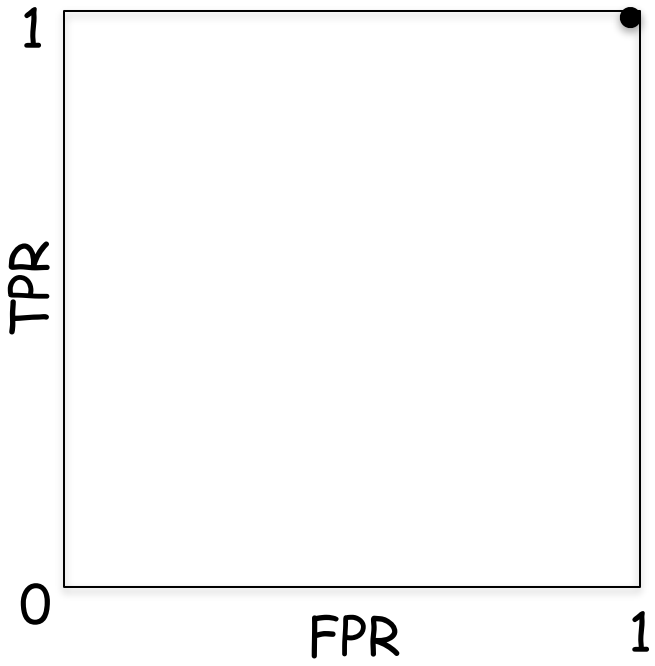
Area Under the Curve (AUC)

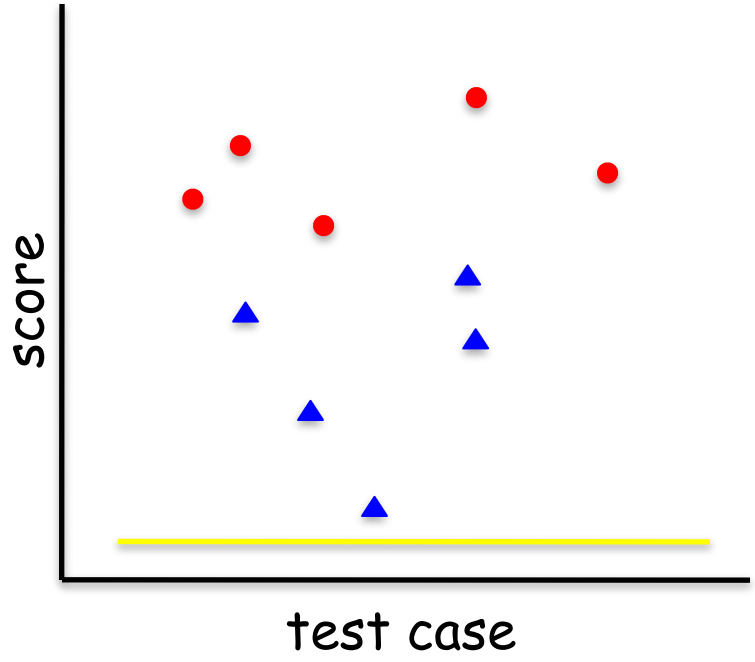
- Area under ROC curve quantifies success
  - 0.5 — like flipping coin
  - 1.0 — ideal detection
- AUC of ROC curve
  - Enables us to compare different techniques
  - And no need to worry about threshold



ROC Curve

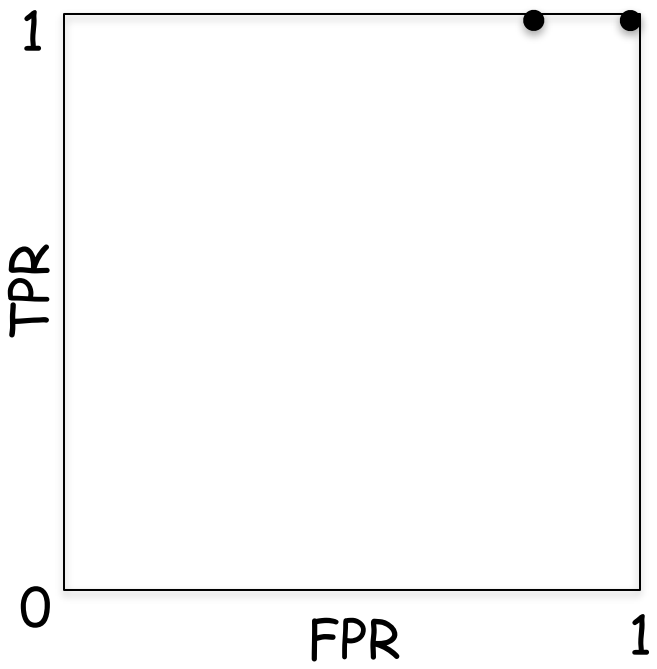
- Another example
  - TPR = 1.0
  - FPR = 1.0 – TNR
    = 1.0 – 0.0 = 1.0

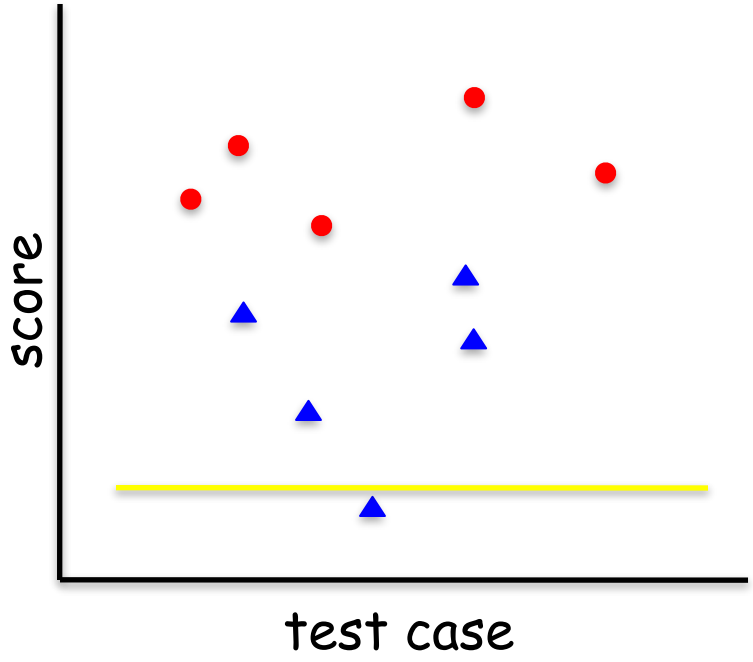




ROC Curve

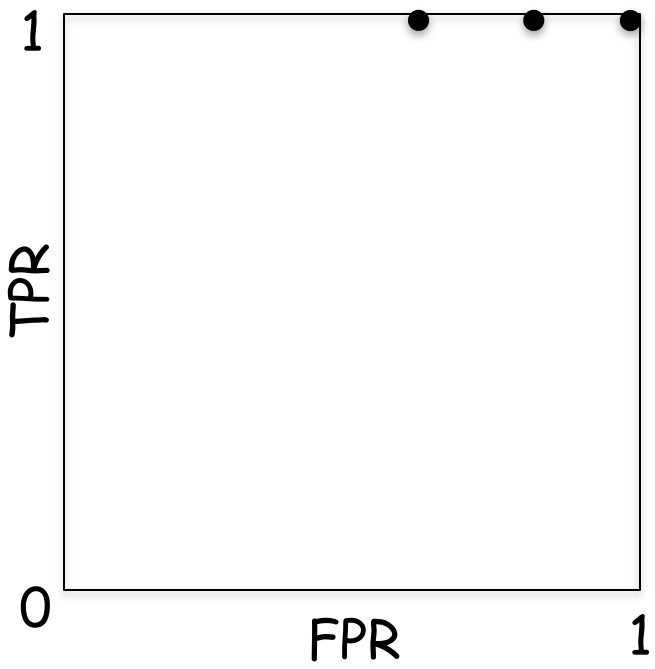
- Another example
  - TPR = 1.0
  - FPR = 1.0 – TNR
    = 1.0 – 0.2 = 0.8

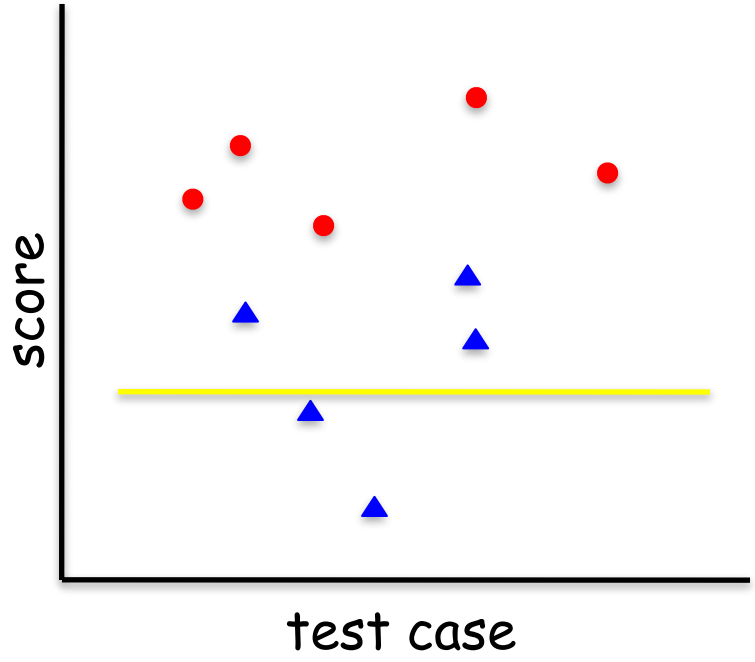




ROC Curve

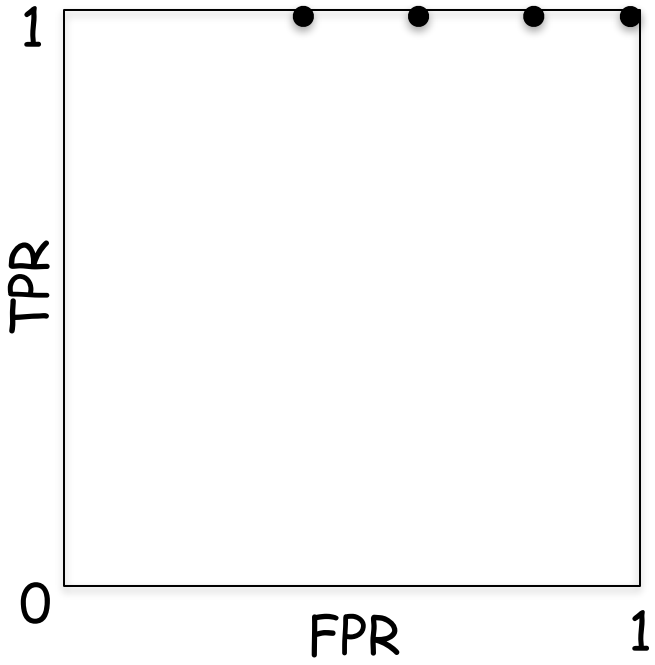
- Another example
  - TPR = 1.0
  - FPR = 1.0 – TNR
    = 1.0 – 0.4 = 0.6

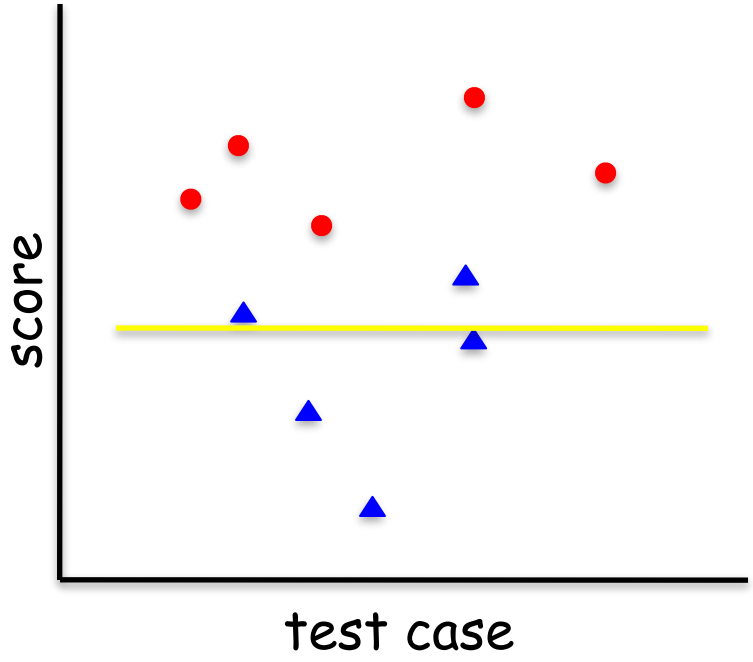




ROC Curve

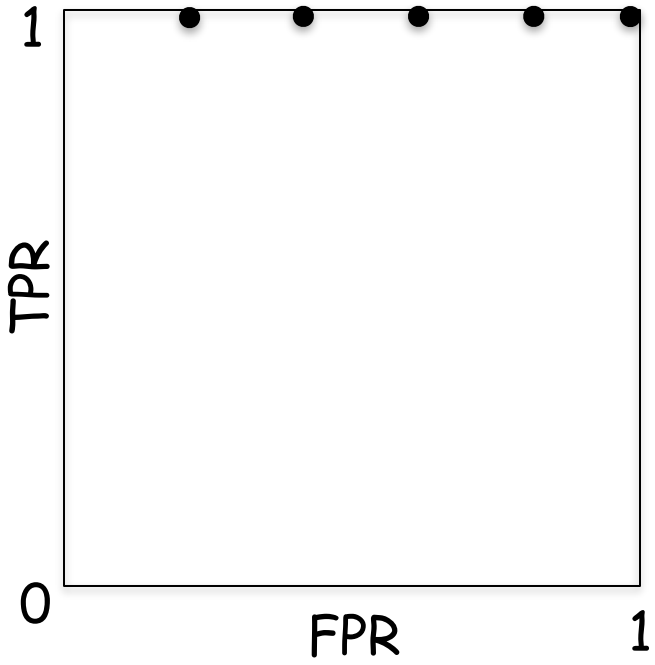
- Another example
  - TPR = 1.0
  - FPR = 1.0 – TNR
    = 1.0 – 0.6 = 0.4

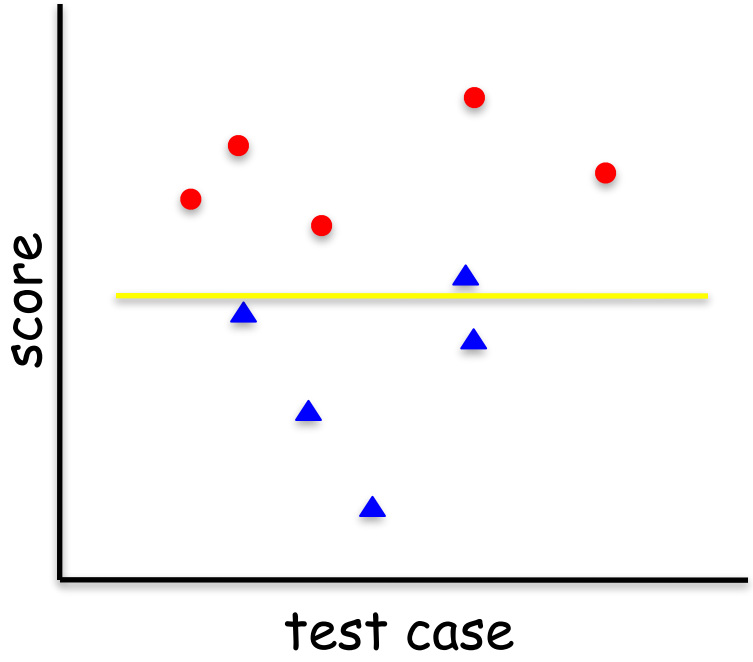




ROC Curve

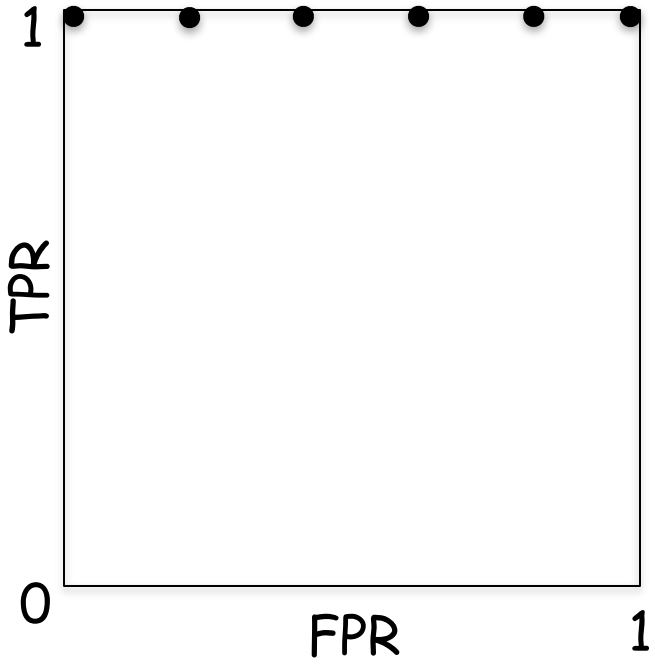
- Another example
  - TPR = 1.0
  - FPR = 1.0 – TNR
    = 1.0 – 0.8 = 0.2

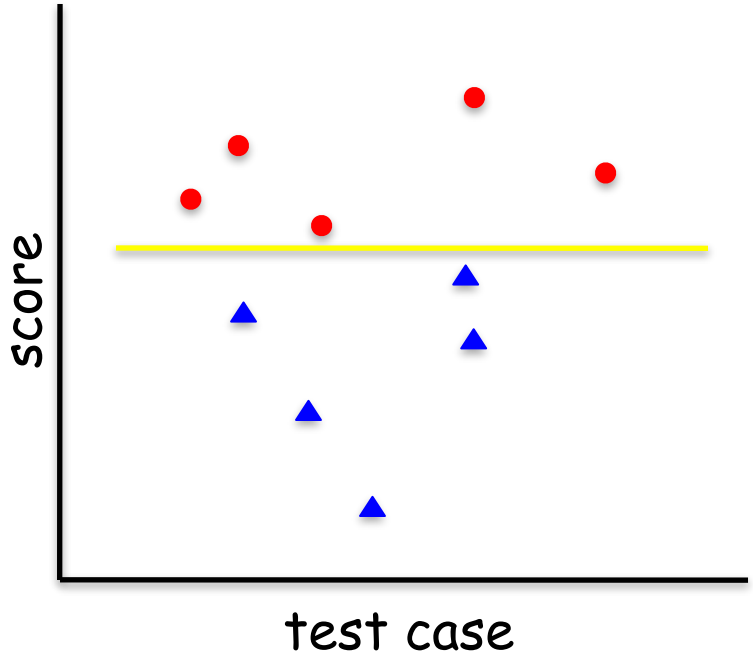




ROC Curve

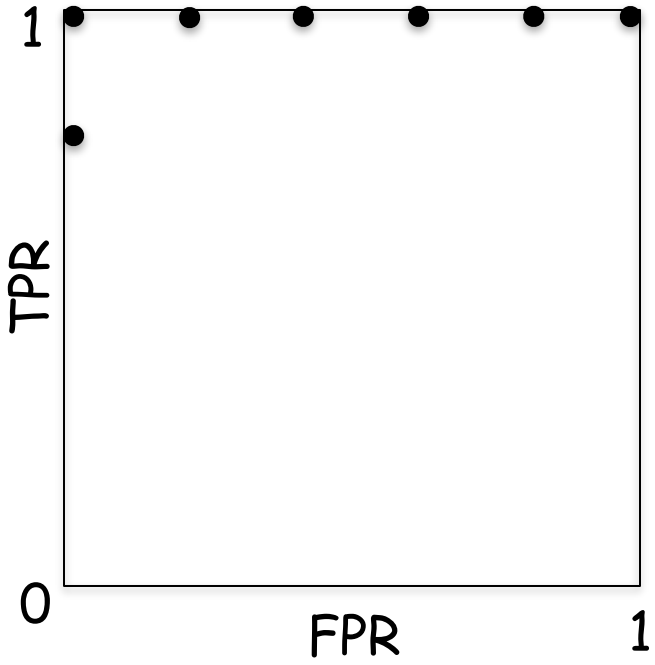
- Another example
  - TPR = 1.0
  - FPR = 1.0 – TNR
    = 1.0 – 1.0 = 0.0

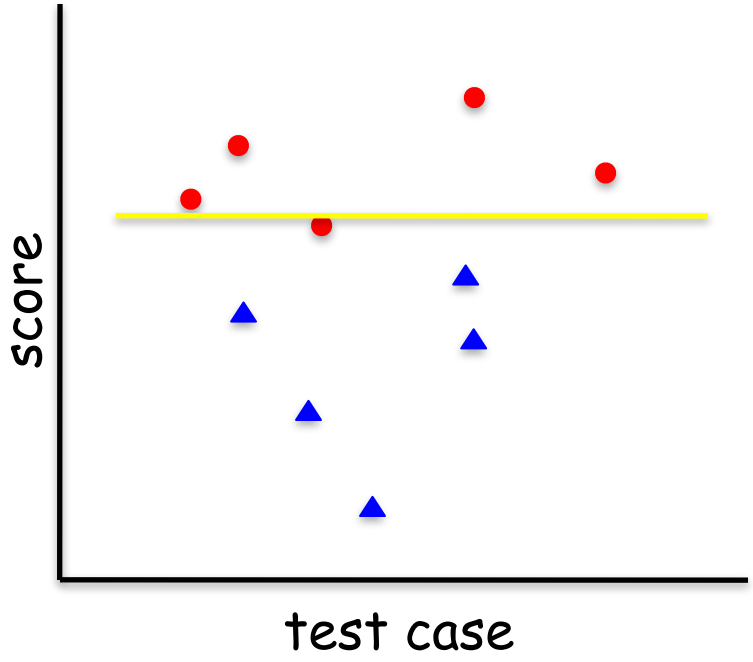




ROC Curve

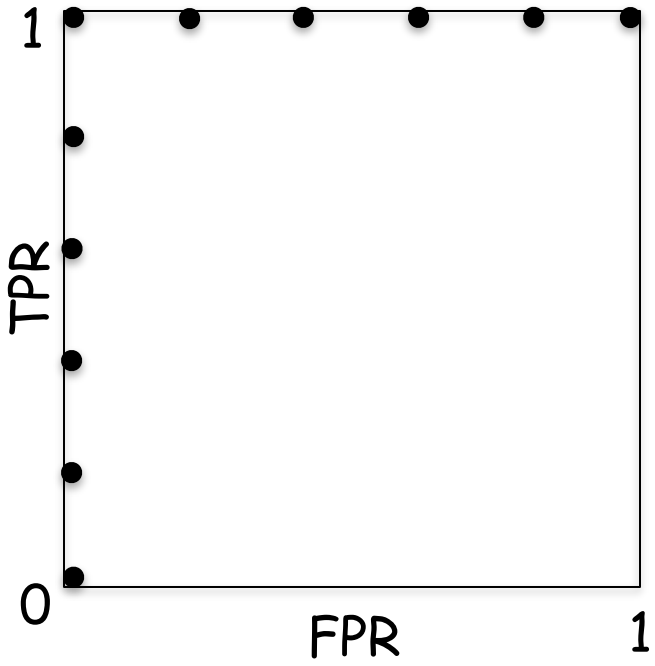
- Another example
  - TPR = 0.8
  - FPR = 1.0 – TNR
    = 1.0 – 1.0 = 0.0

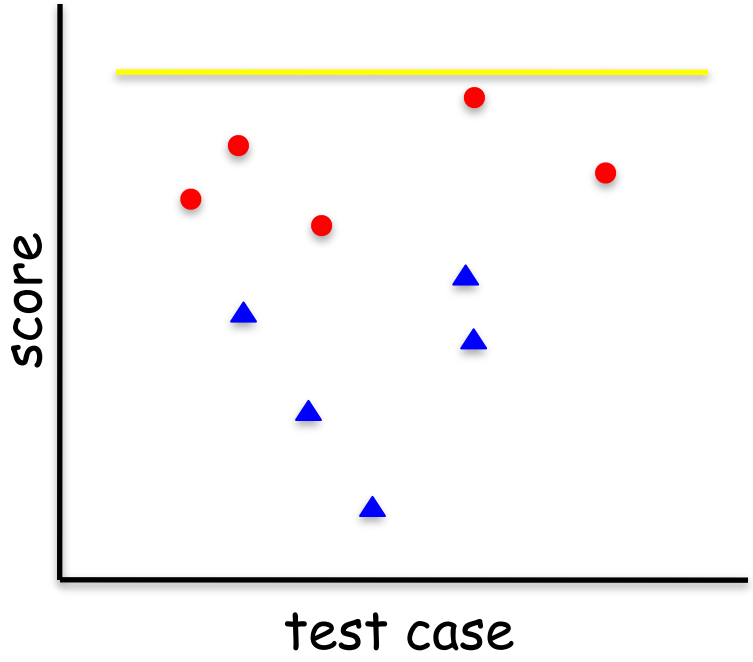




ROC Curve

- Another example
  - TPR = 0.0
  - FPR = 1.0 – TNR
    = 1.0 – 1.0 = 0.0





- What is the difference between the AUC and the Accuracy?
- Just looking at a ROC, can you tell which is the best threshold?