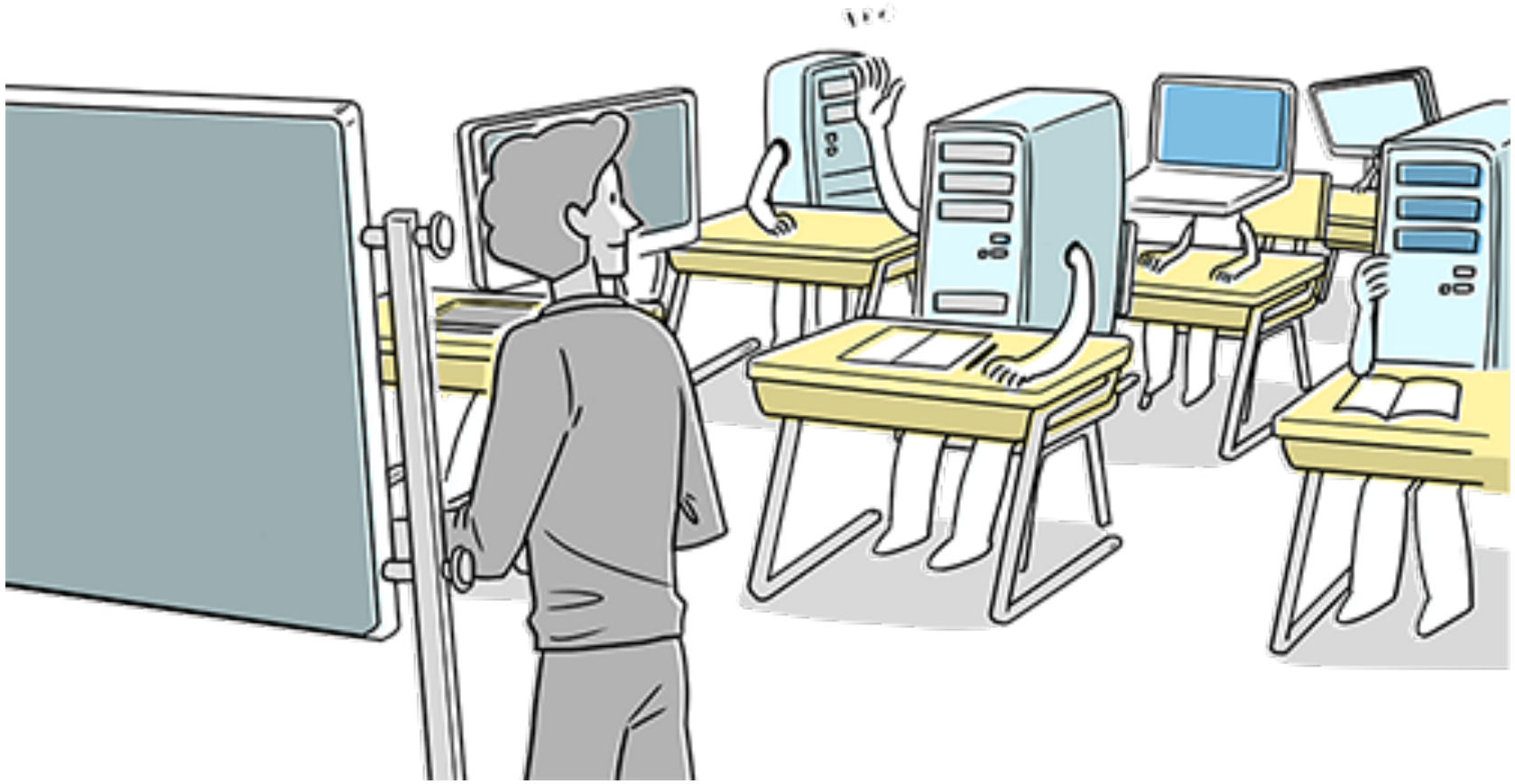


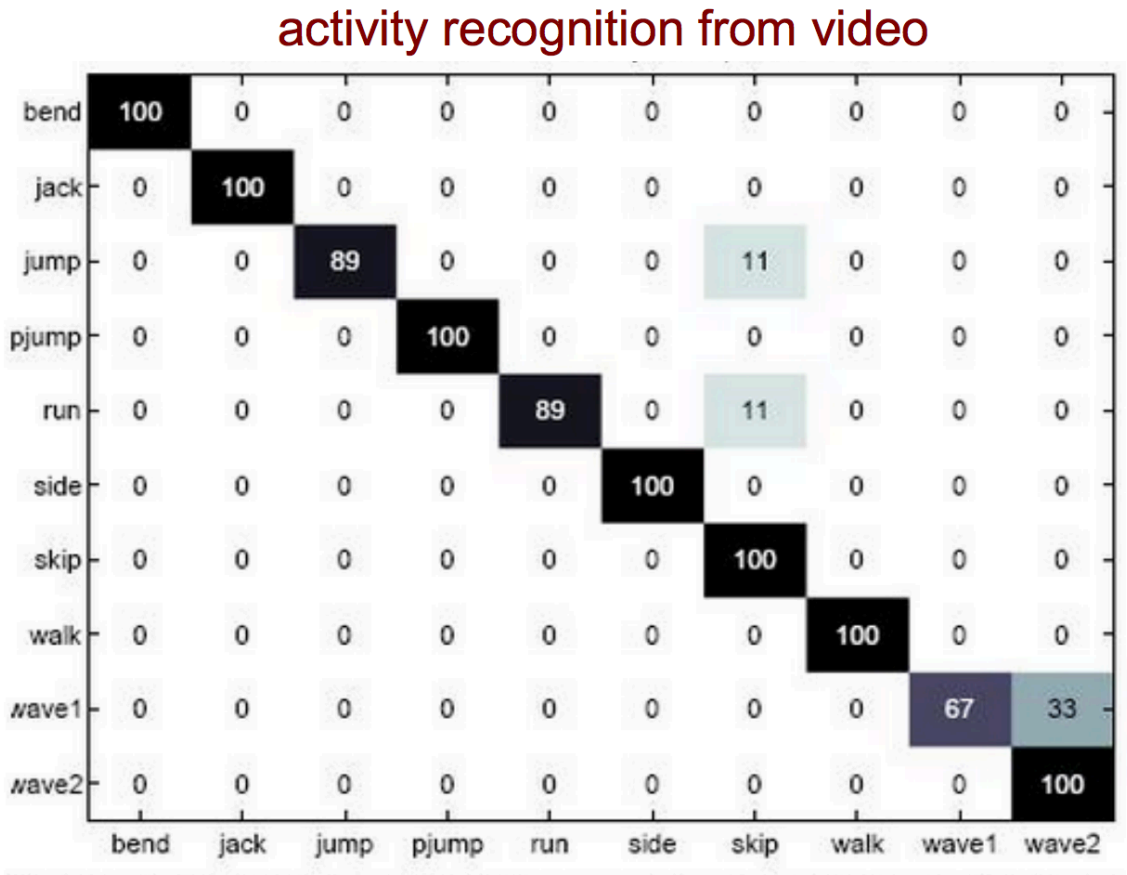
EVALUATION METRICS

Metrics for classification task

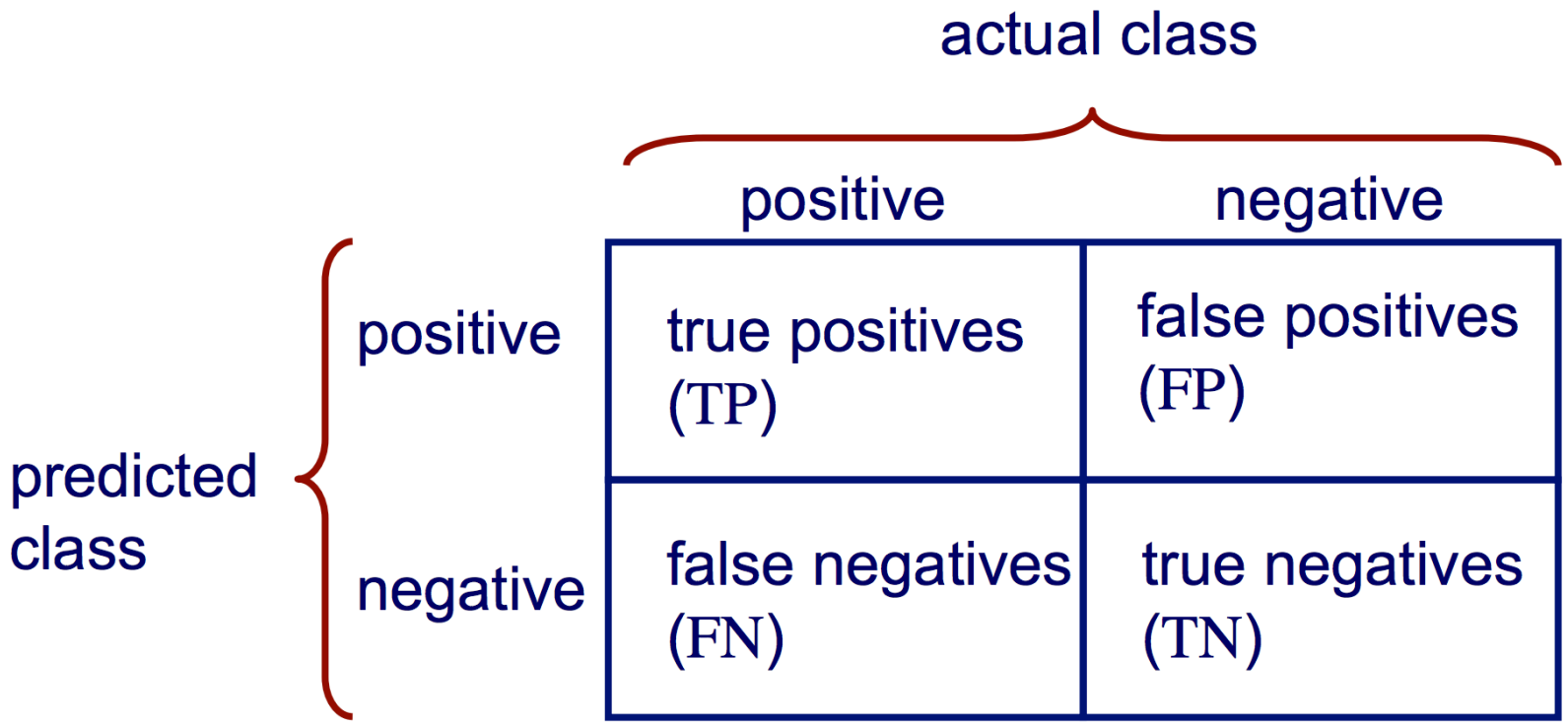
- Confusion Matrix
- Accuracy
- Precision/Recall
- ROC Curve and AUC

Confusion matrix

Understanding what types
of mistakes a learned model makes



Confusion matrix for 2-class problems



$$\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

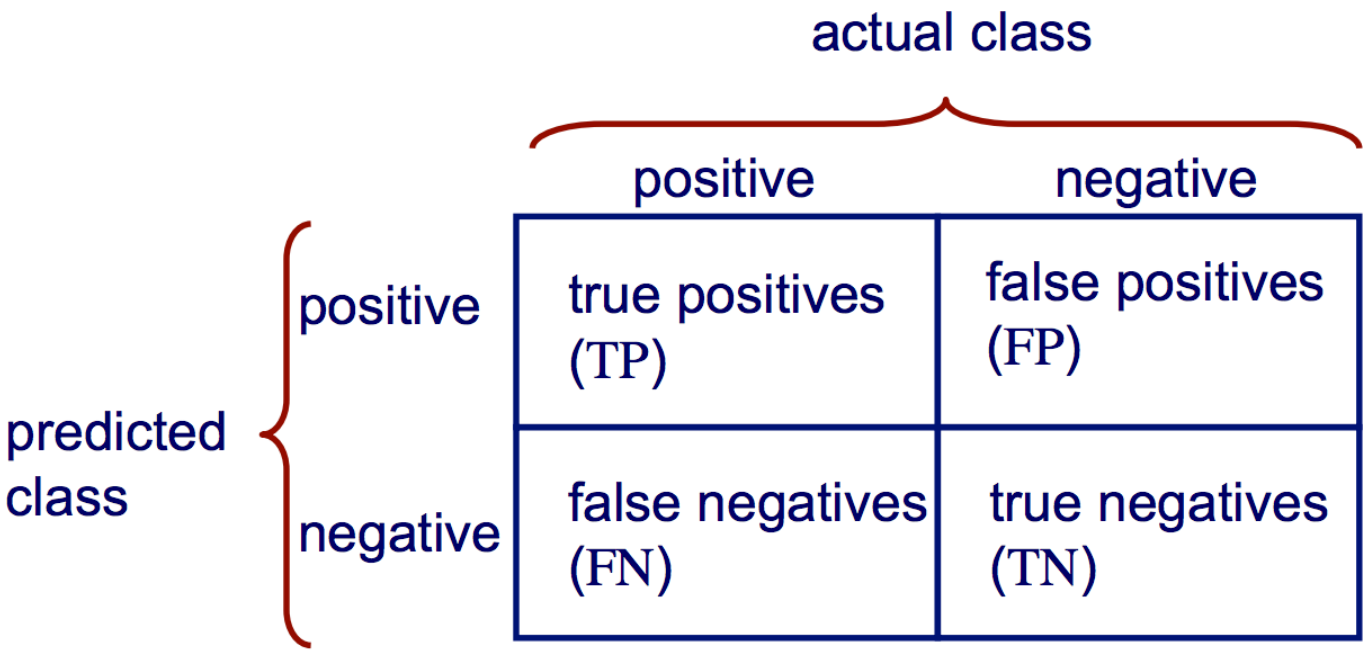
Is accuracy an adequate measure of predictive performance?

Accuracy may not be useful measure in cases where

- There is a large class skew
  - Is 98% accuracy good if 97% of the instances are negative?

- There are differential misclassification costs – say, getting a positive wrong costs more than getting a negative wrong
  - Consider a medical domain in which a false positive results in an extraneous test but a false negative results in a failure to treat a disease

Other accuracy metrics



$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

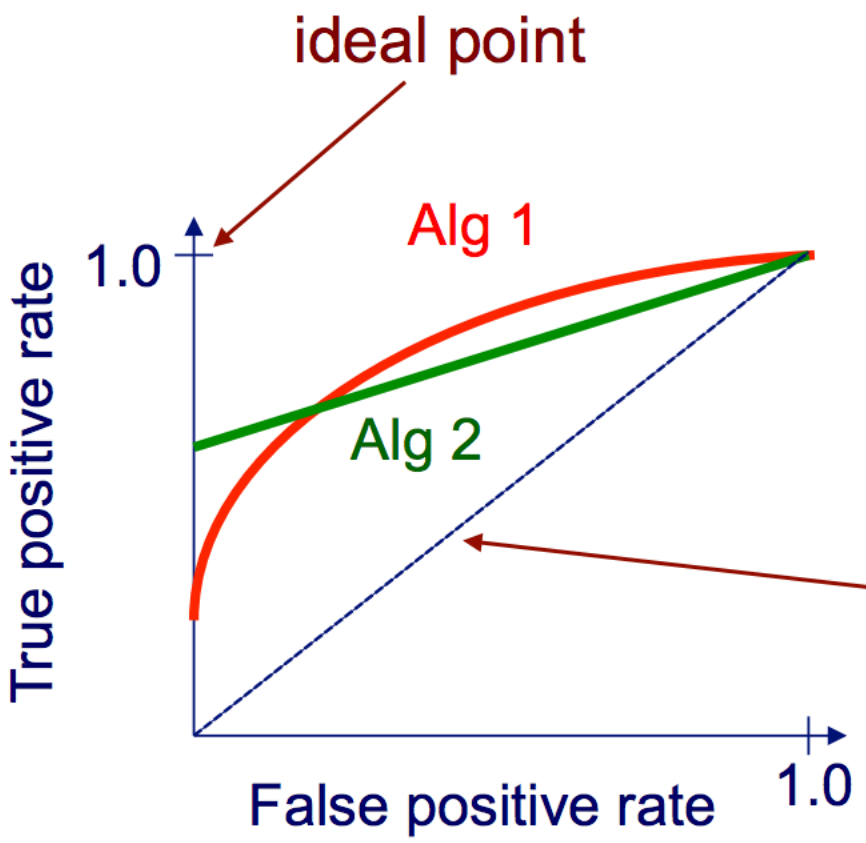
Control the trade-off between precision and recall

$$\text{recall (TP rate)} = \frac{TP}{\text{actual pos}} = \frac{TP}{TP + FN}$$

$$\text{precision} = \frac{TP}{\text{predicted pos}} = \frac{TP}{TP + FP}$$

ROC Curve

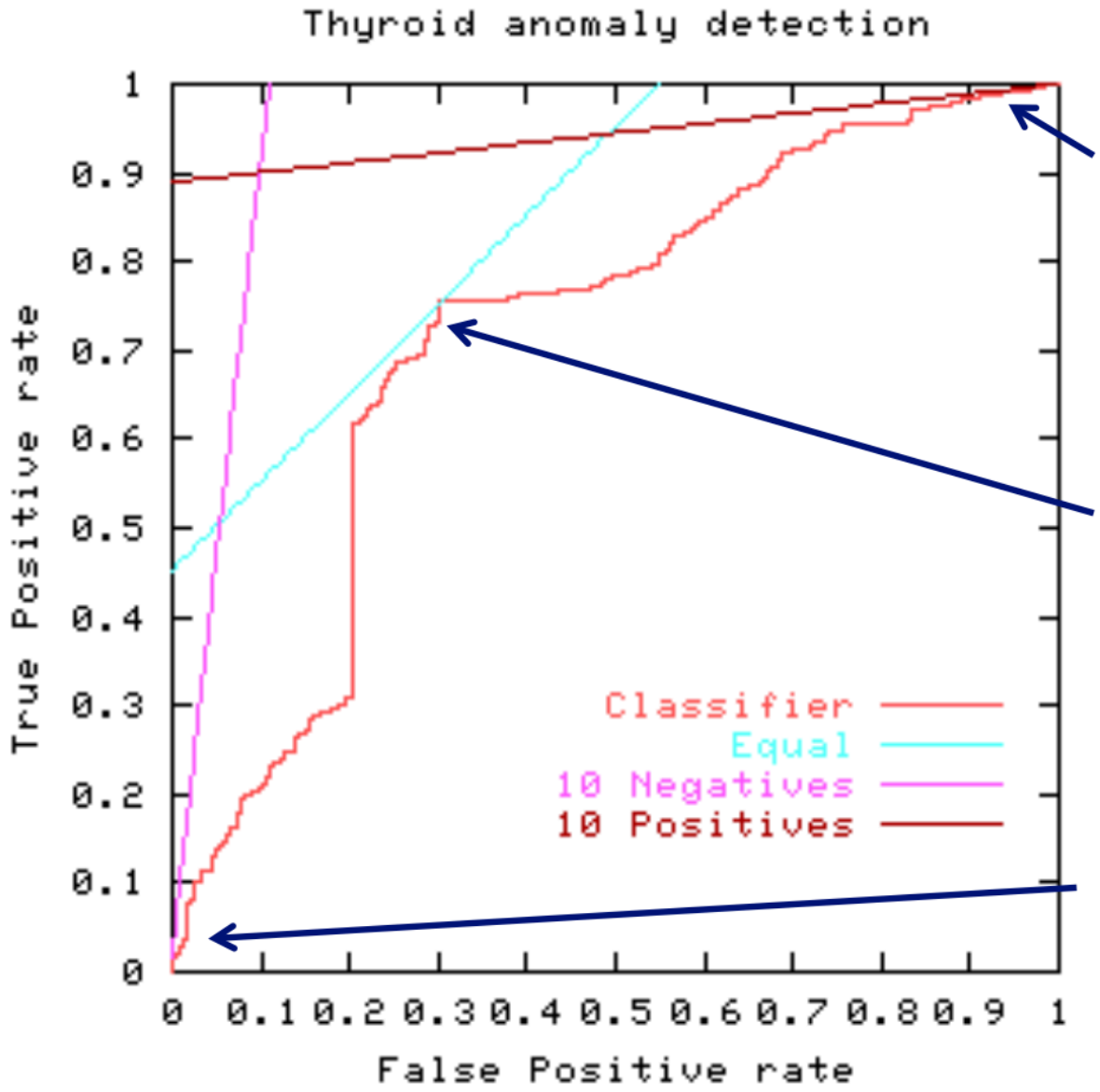
A *Receiver Operating Characteristic* (*ROC*) curve plots the TP-rate vs. the FP-rate as a threshold on the confidence of an instance being positive is varied



Different methods can work better in different parts of ROC space. This depends on cost of false + vs. false -

expected curve for random guessing

ROC curves and misclassification costs

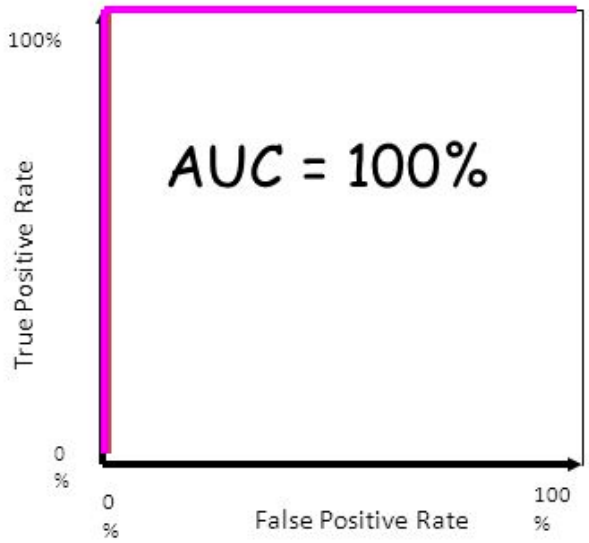


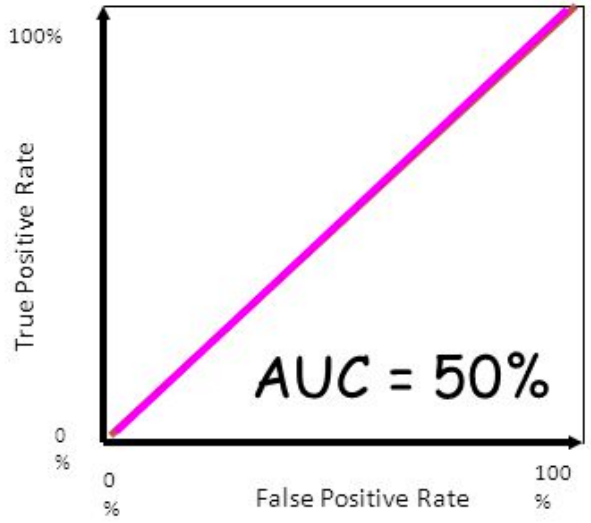
best operating point when FN costs 10× FP

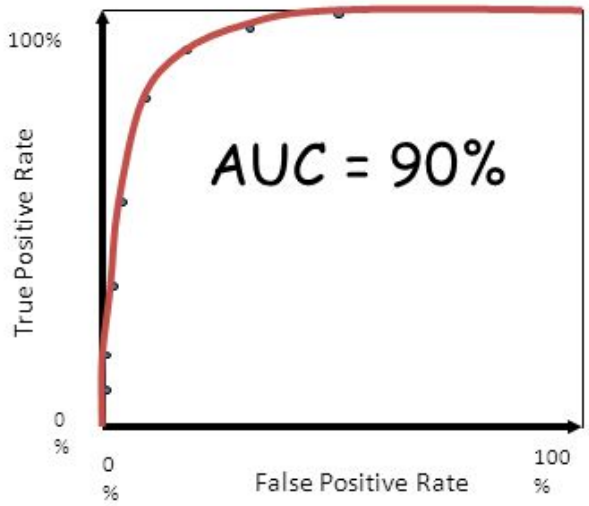
best operating point when cost of misclassifying positives and negatives is equal

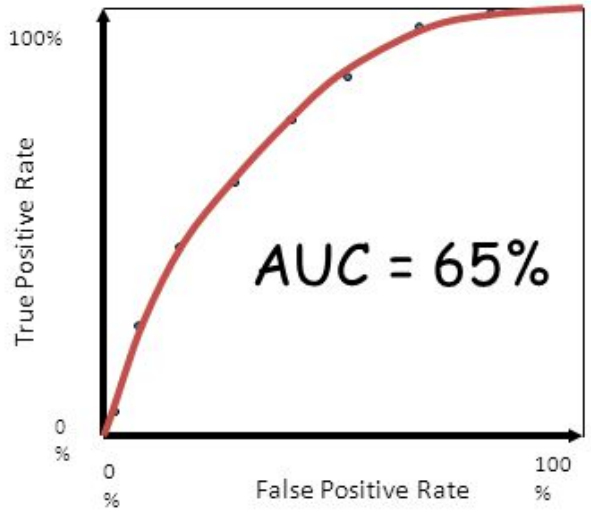
best operating point when FP costs 10× FN

AUC for ROC Curve









It doesn't care about absolute values, it only cares about ranking

Get perfect score when:
the predicted score of positive class is higher than negative class in all examples

Metrics for regression task

- Root Mean Square Error (RMSE)
  - Most widely used
  - Emphasize bigger deviations

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

- Mean Absolute Error (MAE)
  - Easiest to interpret

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$