

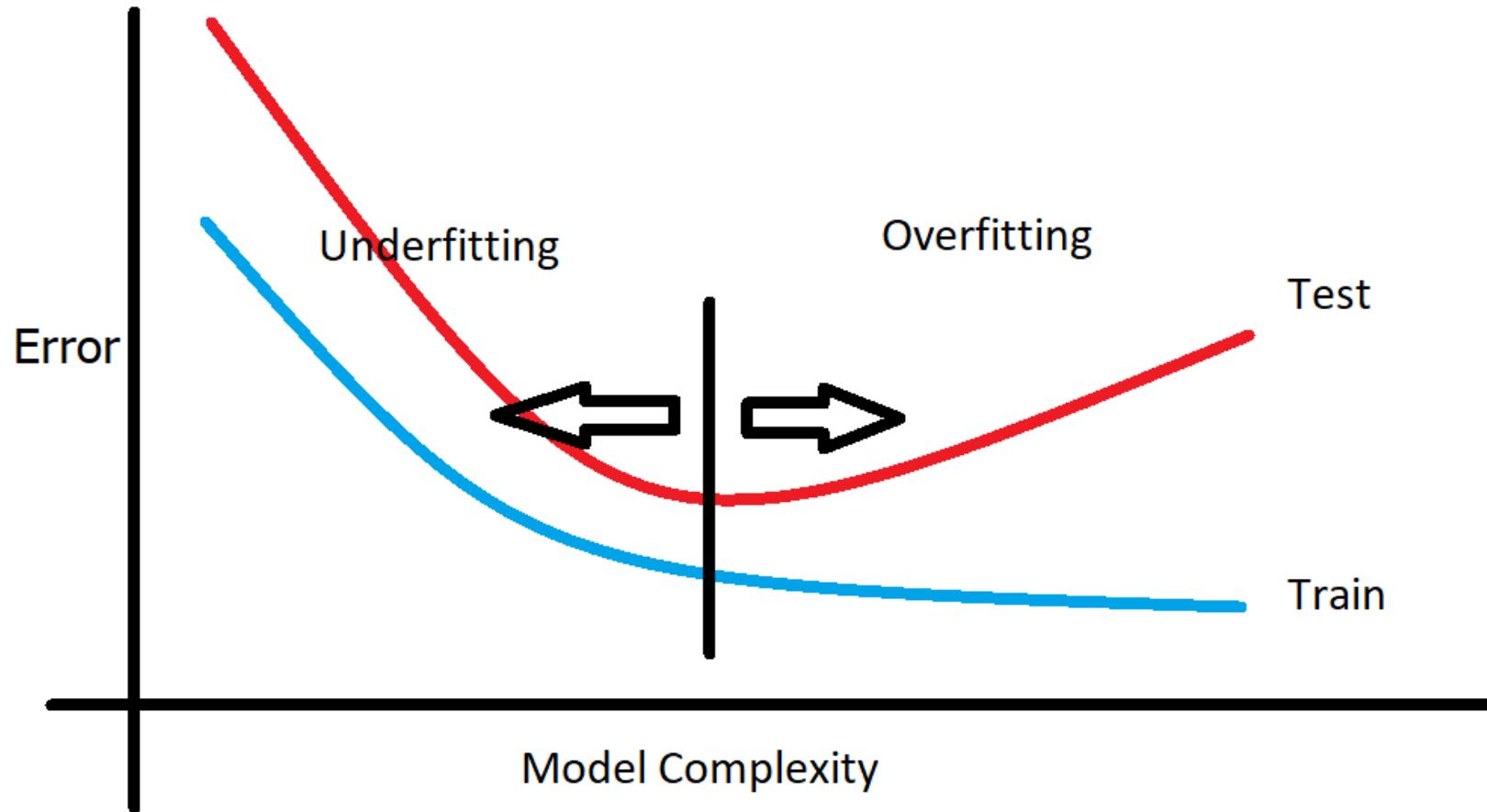
# Evaluation of ML Models

Dr. Uzair Ahmad

# Agenda

- Estimating model fit
  - Train-test split
  - Cross Validation
- Bias-Variance Trade-off
- Measures of performance
  - Confusion Matrix
  - Precision, Recall, TPR, FPR
  - ROC Curve

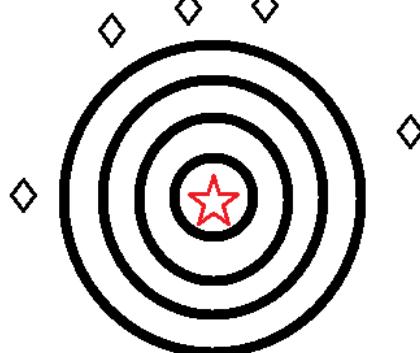
# Training-Test Error



# Bias Variance

1.

High variance  
High Bias



2.

Low variance  
High Bias



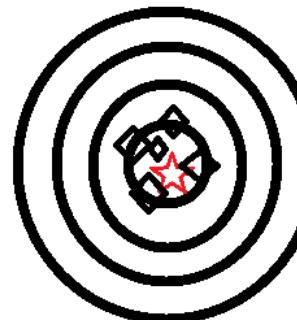
3.

High Variance  
Low Bias



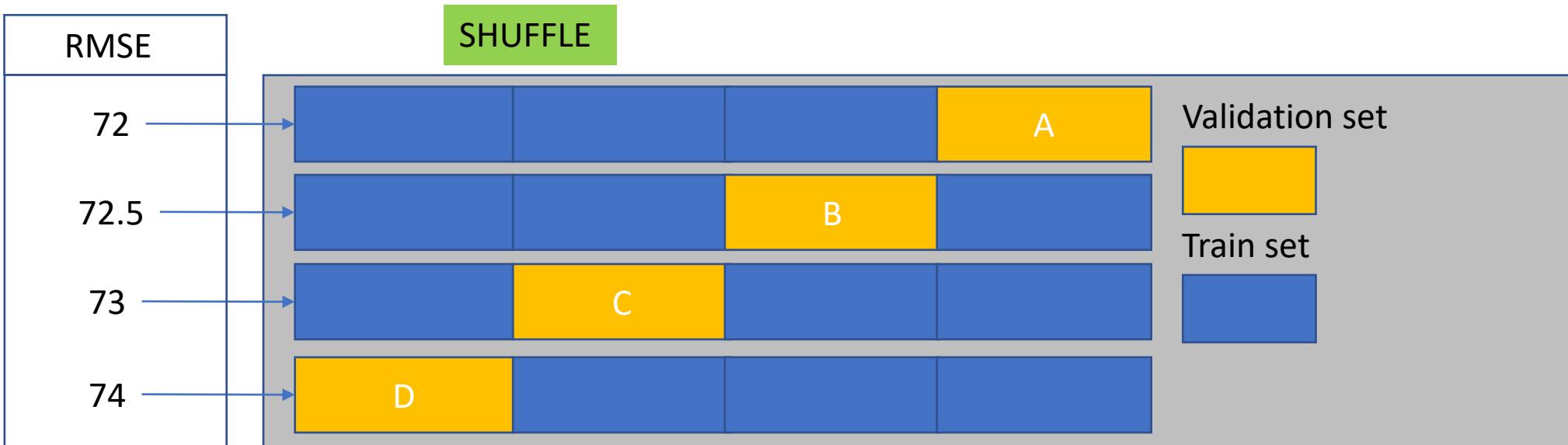
4.

Low variance  
Low Bias



# K-Fold Cross Validation

Ensure Generalizability

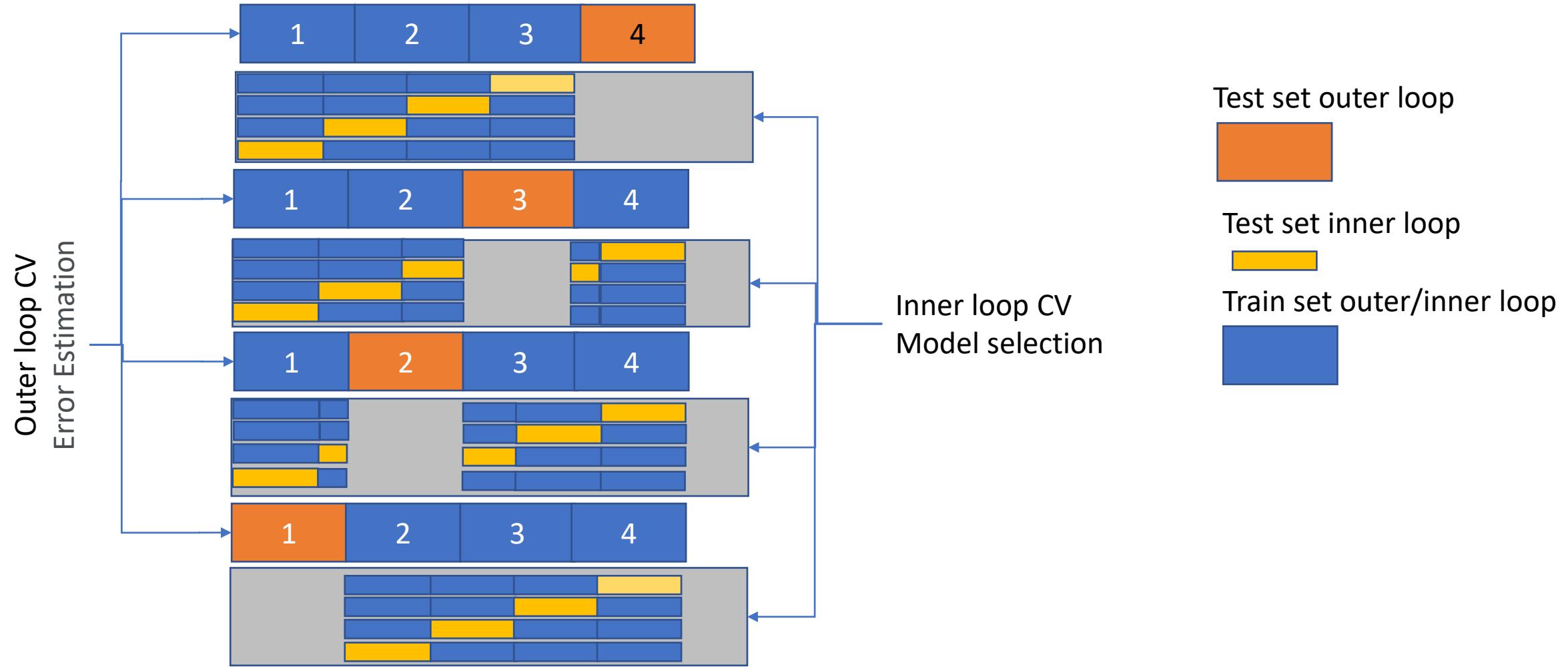


ANOVA: Analysis of Variance → Null Hypothesis: All are same.

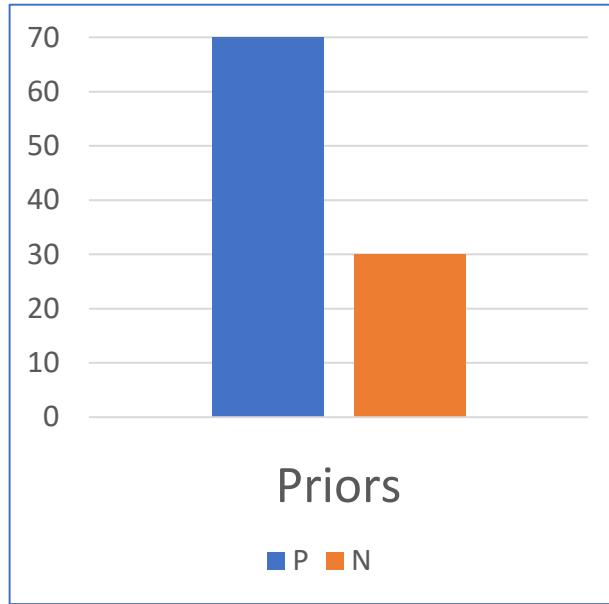
Q. Which one of the four models will be used?

$$\begin{aligned}Y &= m_1x + b_1 \\Y &= m_2x + b_2 \\Y &= (m_1 + m_2)x + (b_1 + b_2)\end{aligned}$$

# Nested CV



# Evaluating Classifiers

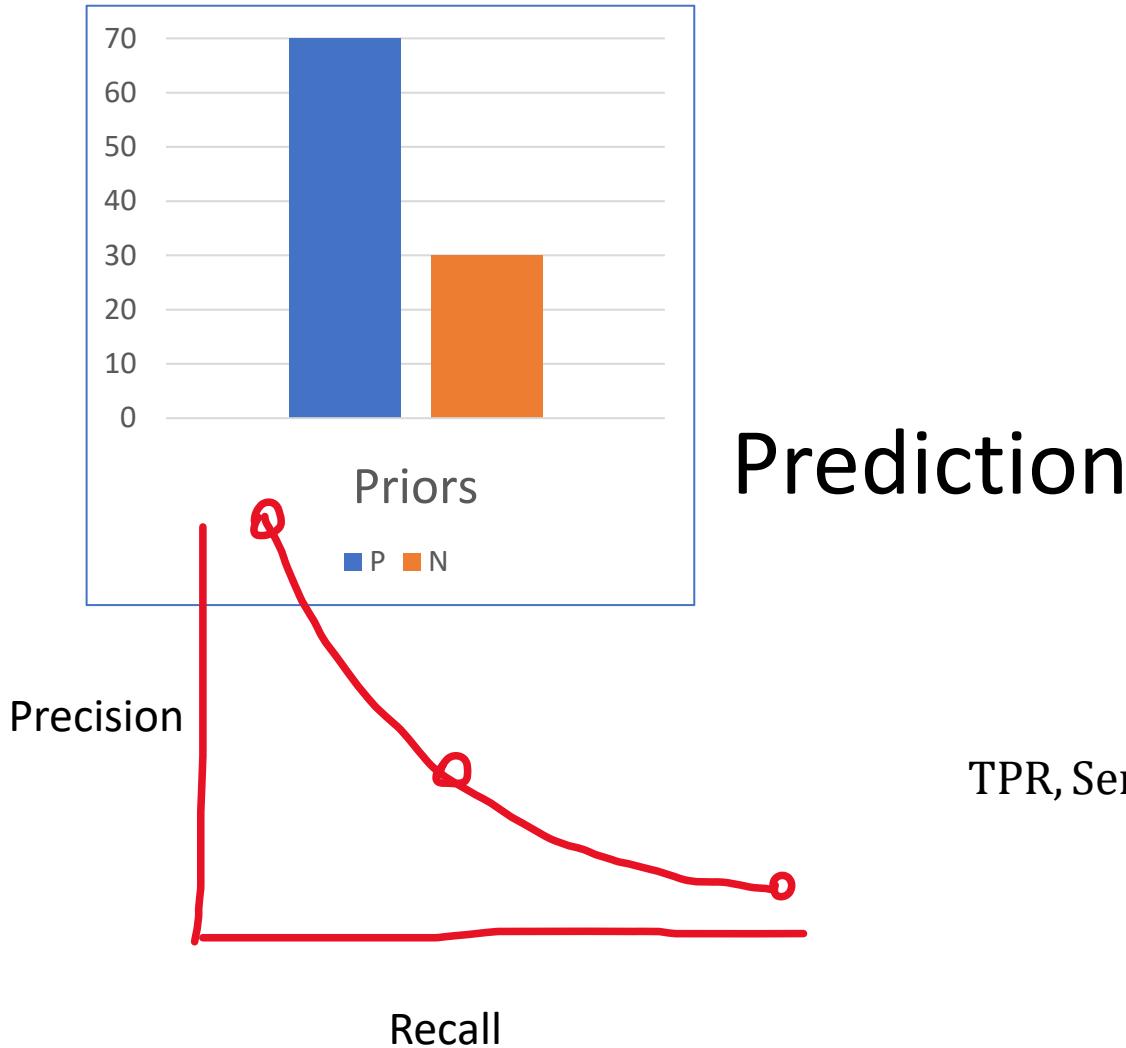


Prediction

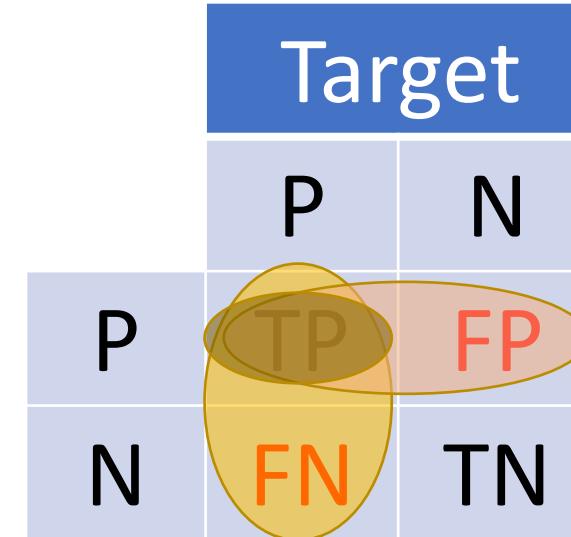
| Target |    |    |    |
|--------|----|----|----|
|        |    | P  | N  |
| P      | TP | FP |    |
|        | N  | FN | TN |

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

# Evaluating Classifiers



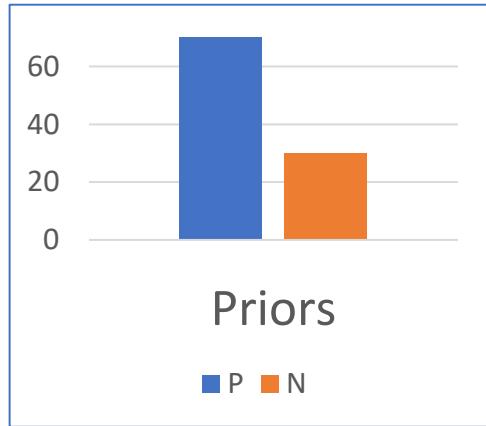
Prediction



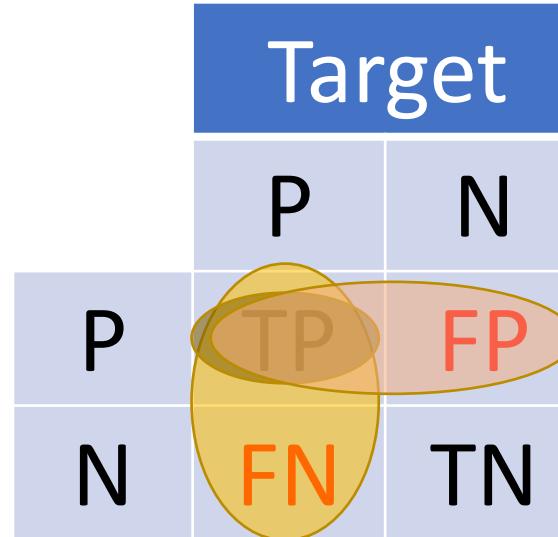
$$\text{TPR, Sensitivity, Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

# Evaluating Classifiers



Prediction

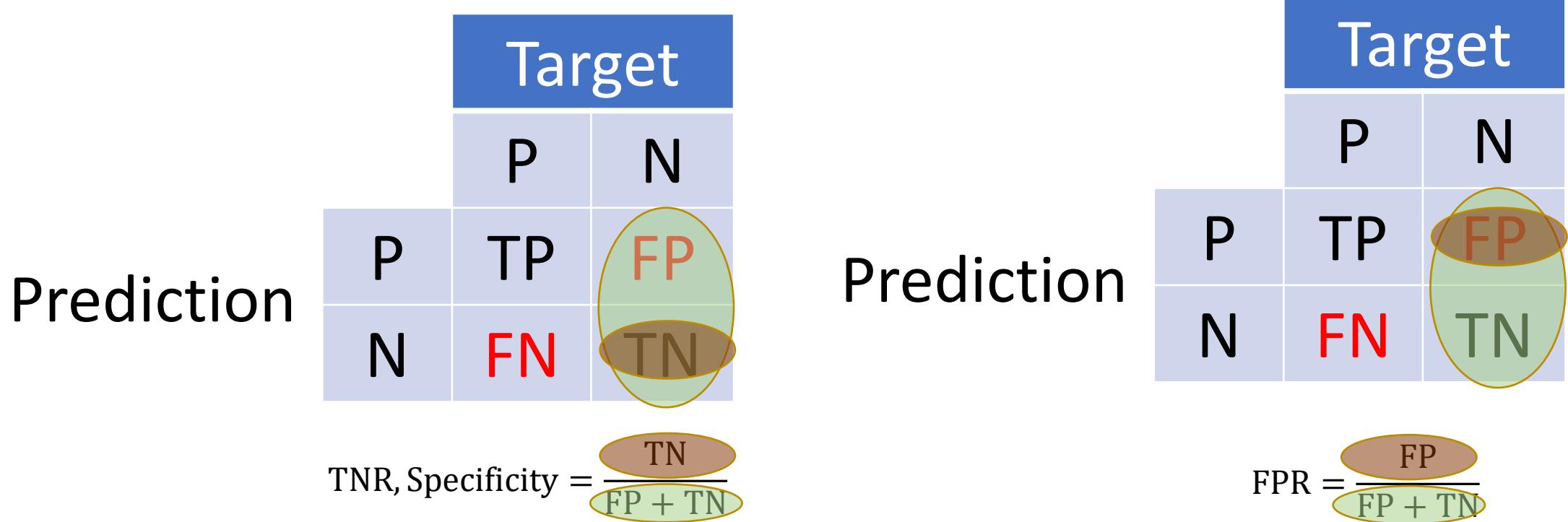


$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$F - \text{Measure} = \frac{2 \text{ Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

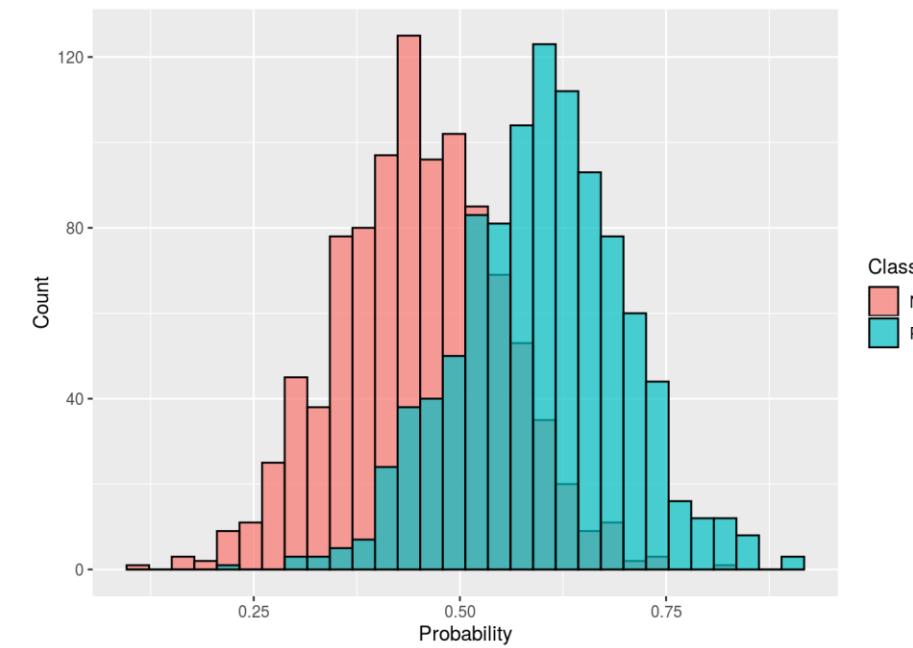
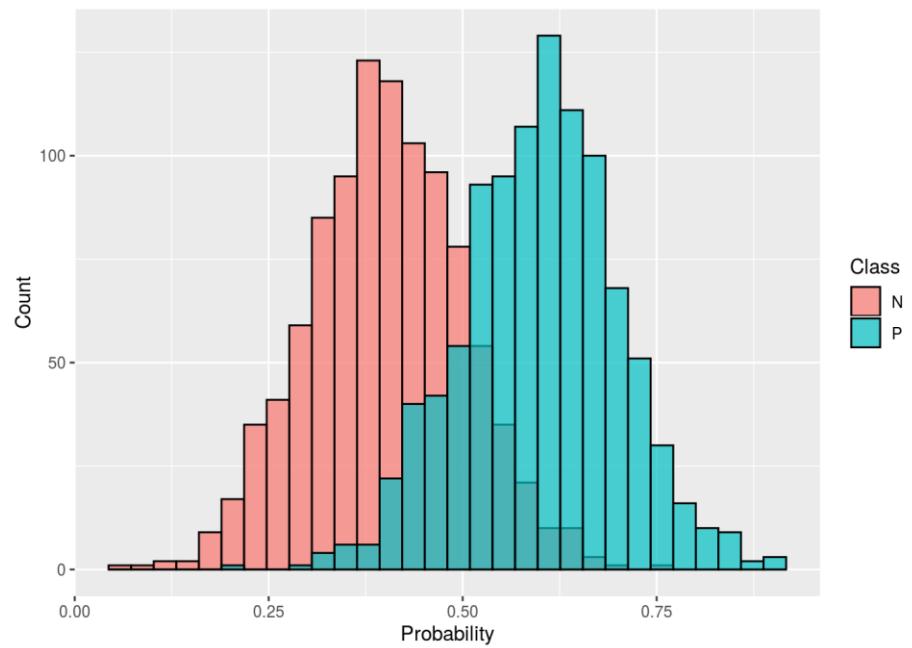
$$\text{TPR, Sensitivity, Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

# Evaluating Classifiers



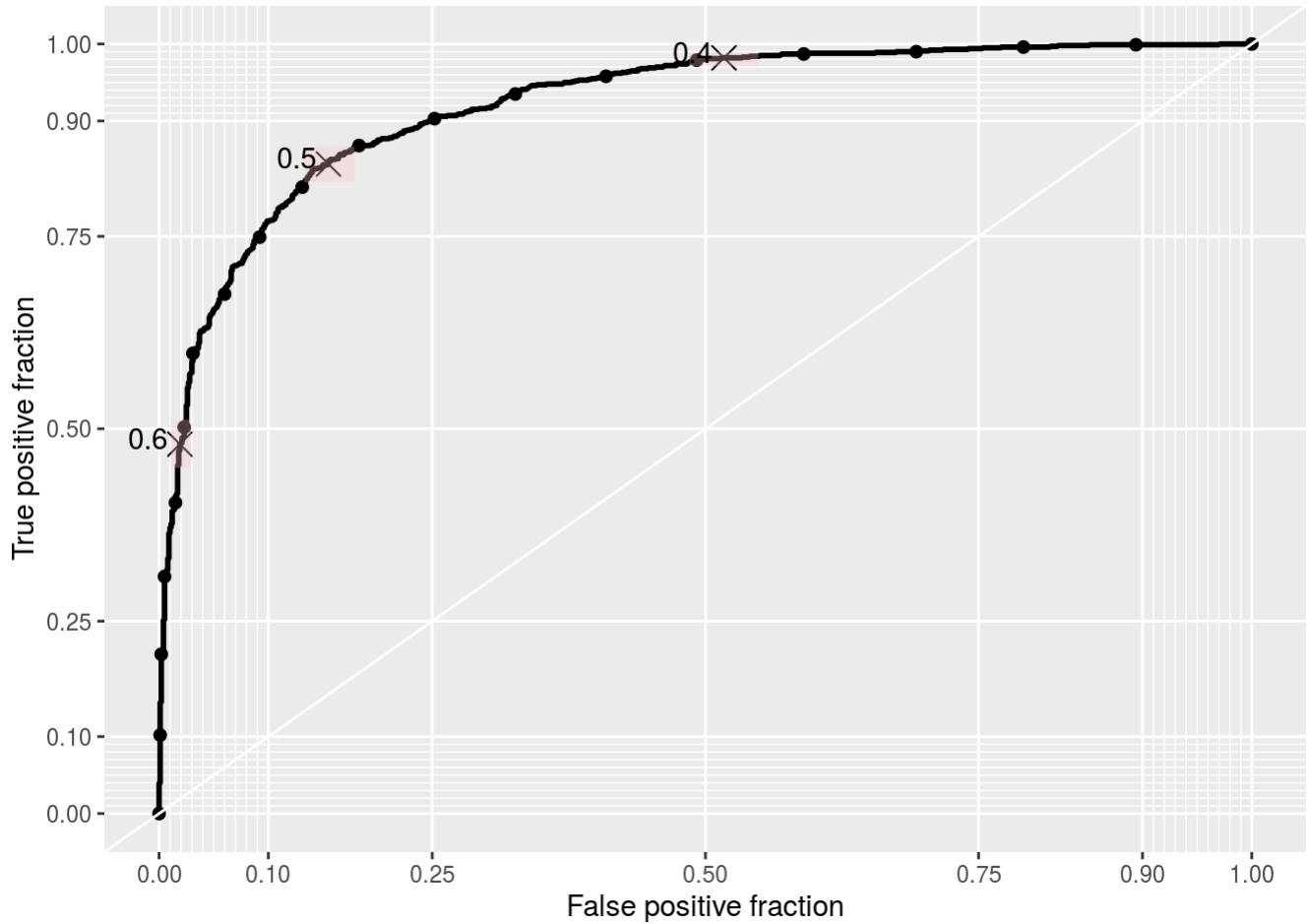
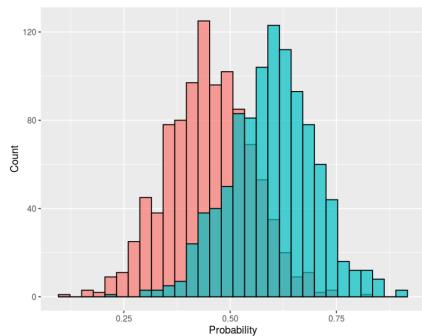
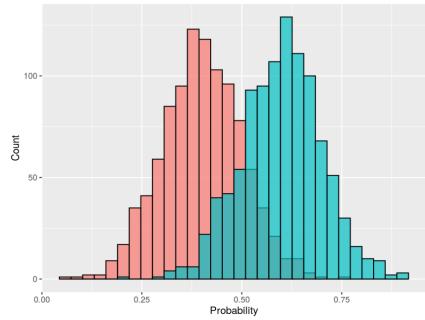
# Evaluating Classifiers

- Threshold free measures



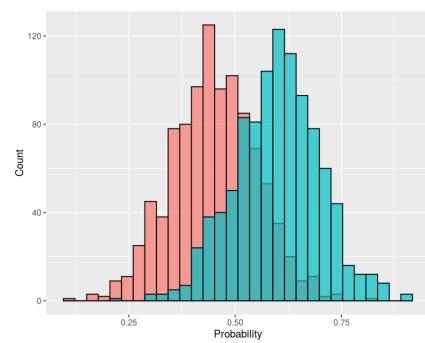
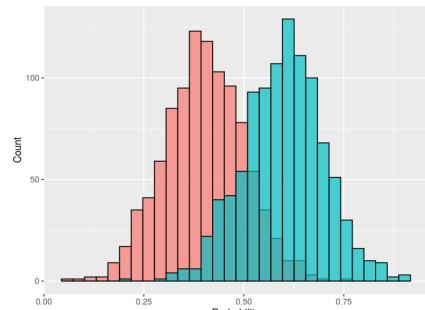
# Evaluating Classifiers

- Threshold free measures

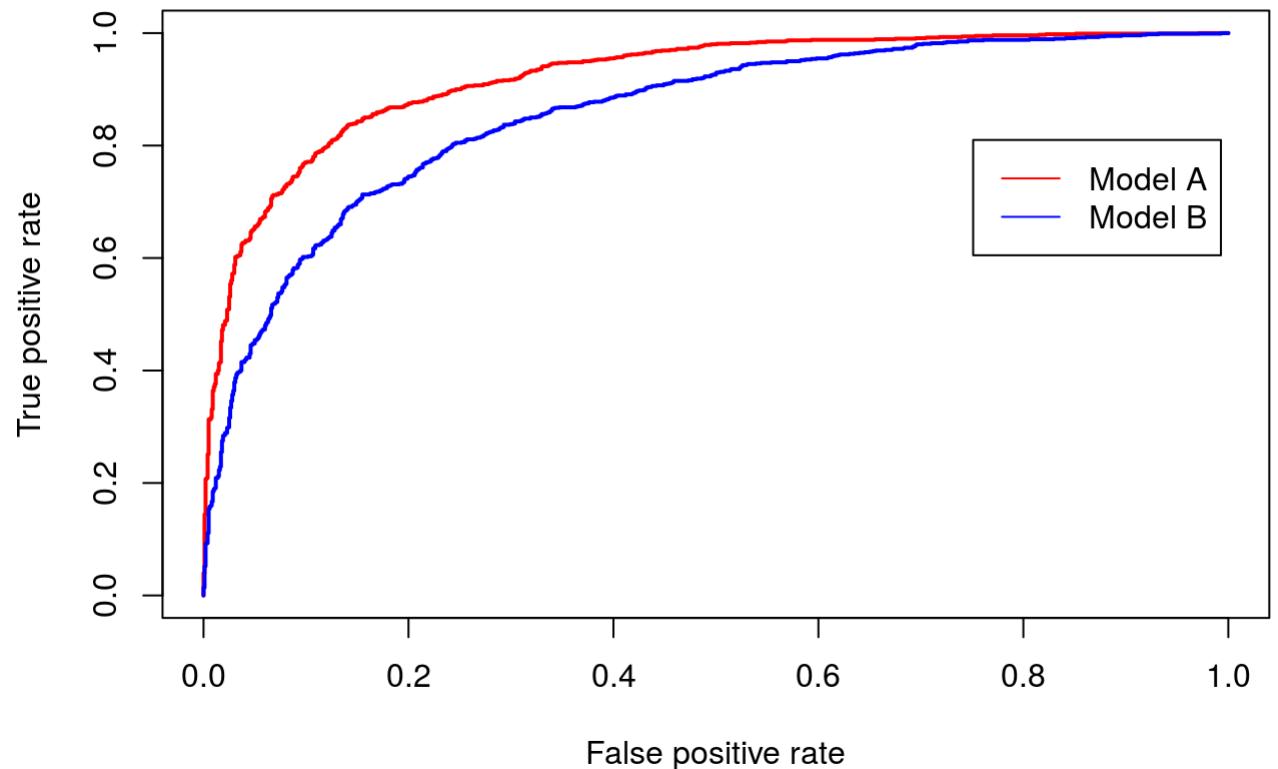


# Evaluating Classifiers

- Threshold free measures

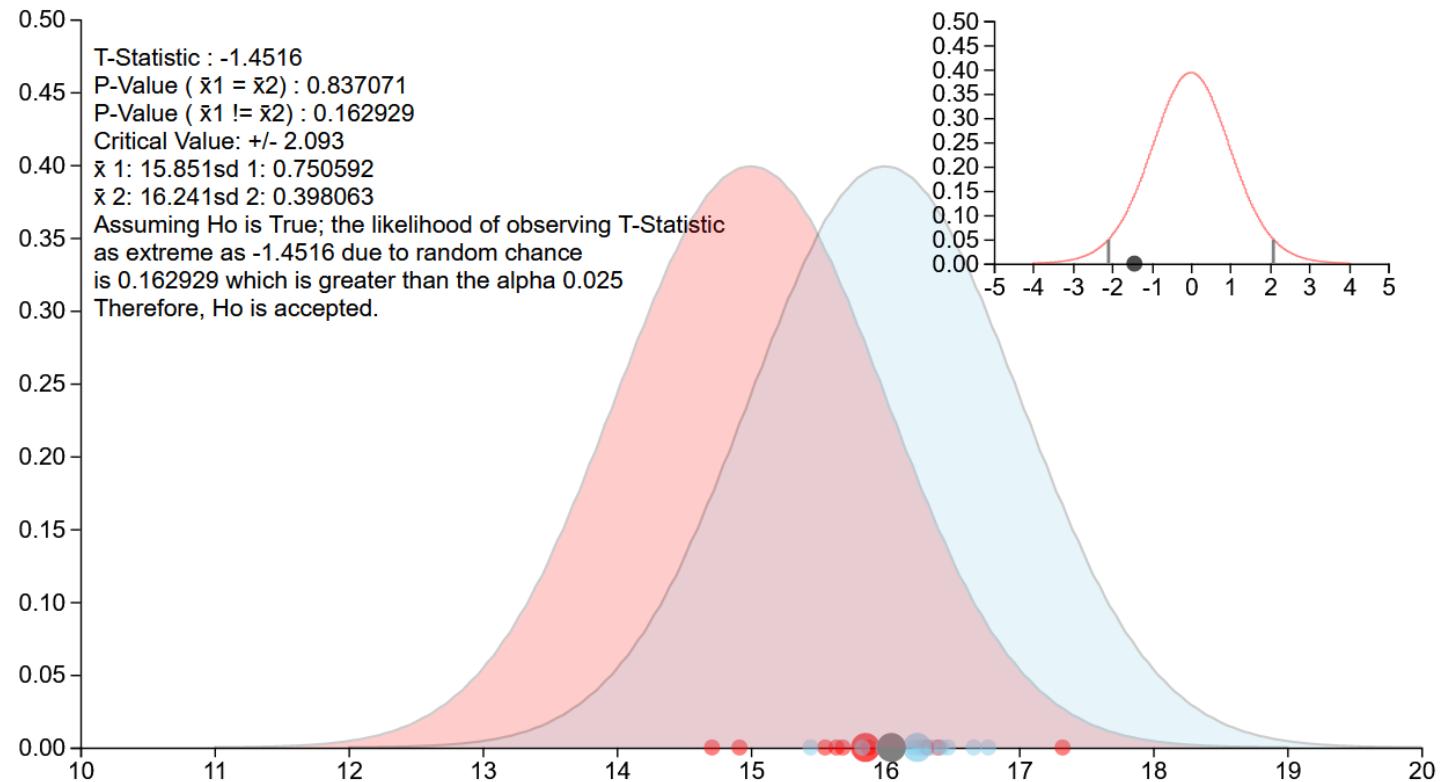


**ROC: Comparing Models**  
**Model A AUC: 0.92, Model B AUC: 0.86**



# Evaluating Classifiers

- Comparing two models
  - Students' T-Test



# Evaluating Classifiers

- Source of variation
  - Algorithm
  - Regularization
  - Feature set
  - Initialization
  - Order of inputs
- Ablation study
  - Statistical significance

$$H_0: C_1^{full} = C_1^{ablated}$$

# Evaluating Classifiers

- Hypothesis Testing
  - Symmetric

$$H_0: C_1 = C_2$$

- Asymmetric

$$H_0: C_1 \leq C_2$$

$$H_0: C_1 \geq C_2$$