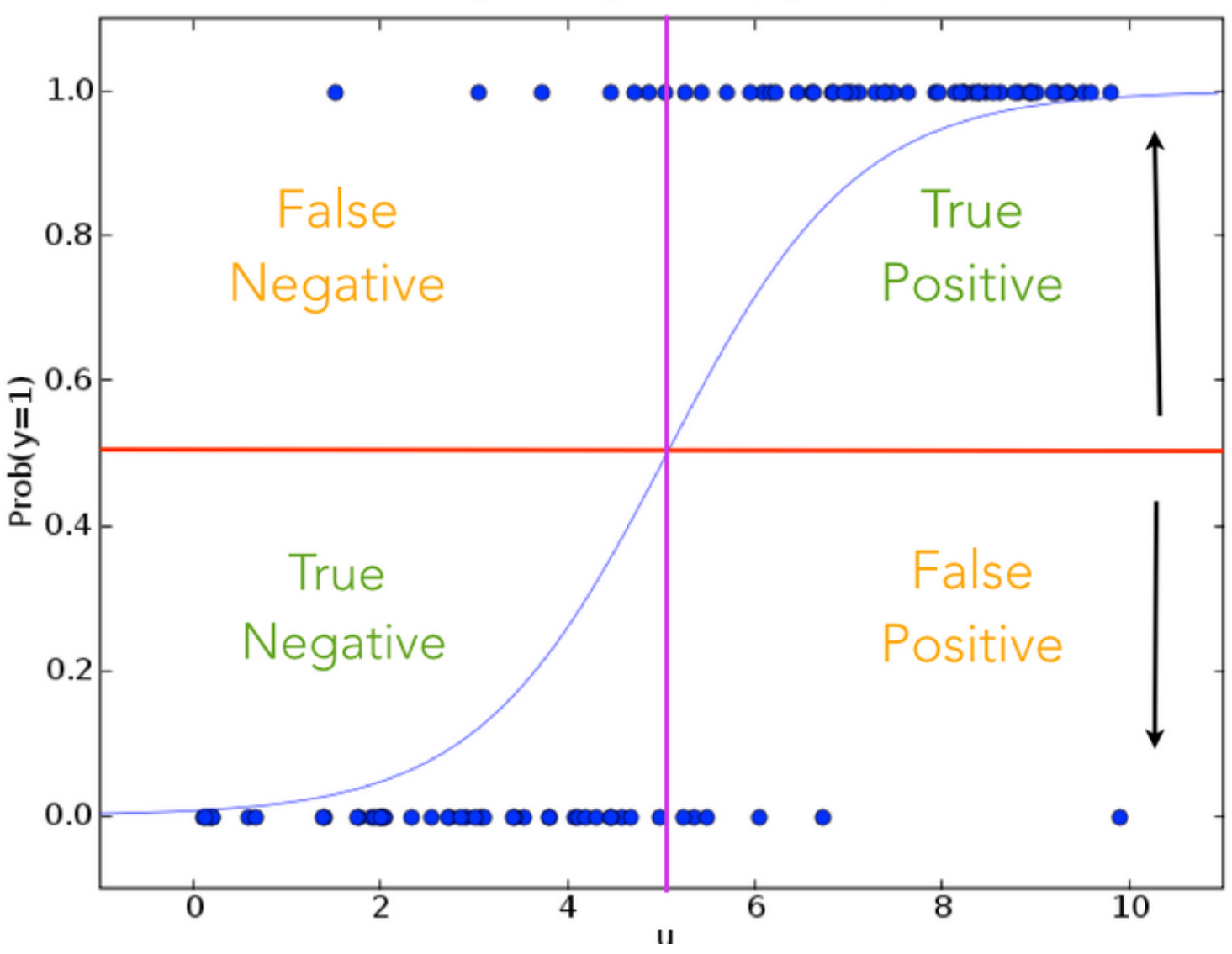
ROC Curves

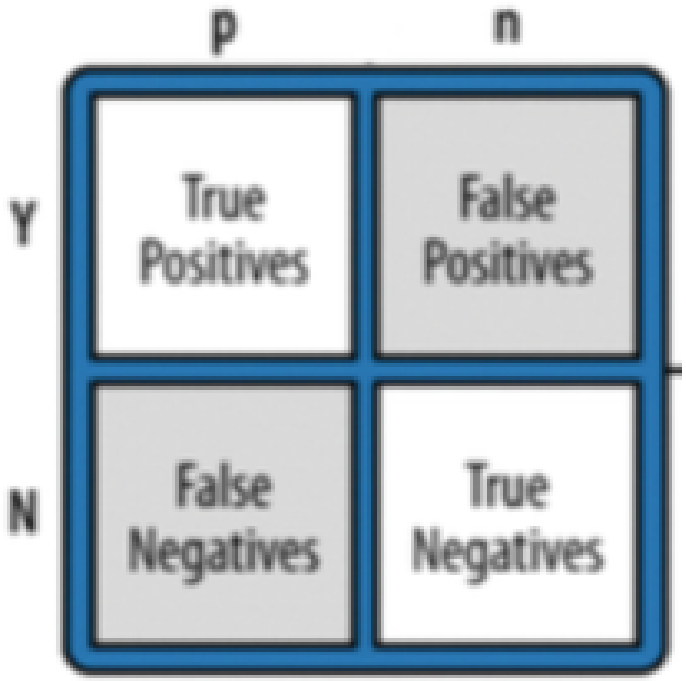
Clayton W. Schupp

Galvanize

Logistic Regression Revisited



Confusion Matrix



- TP rate $= \frac{TP}{P}$
- FP rate $= \frac{FP}{N}$
- Accuracy $= \frac{TP+TN}{P+N}$
- Precision $= \frac{TP}{TP+FP}$
- Sensitivity $=$ recall
- Specificity $= \frac{TN}{FP+TN}$

Building the ROC Curve

For a given model f, each threshold value T gives a point on the ROC Curve

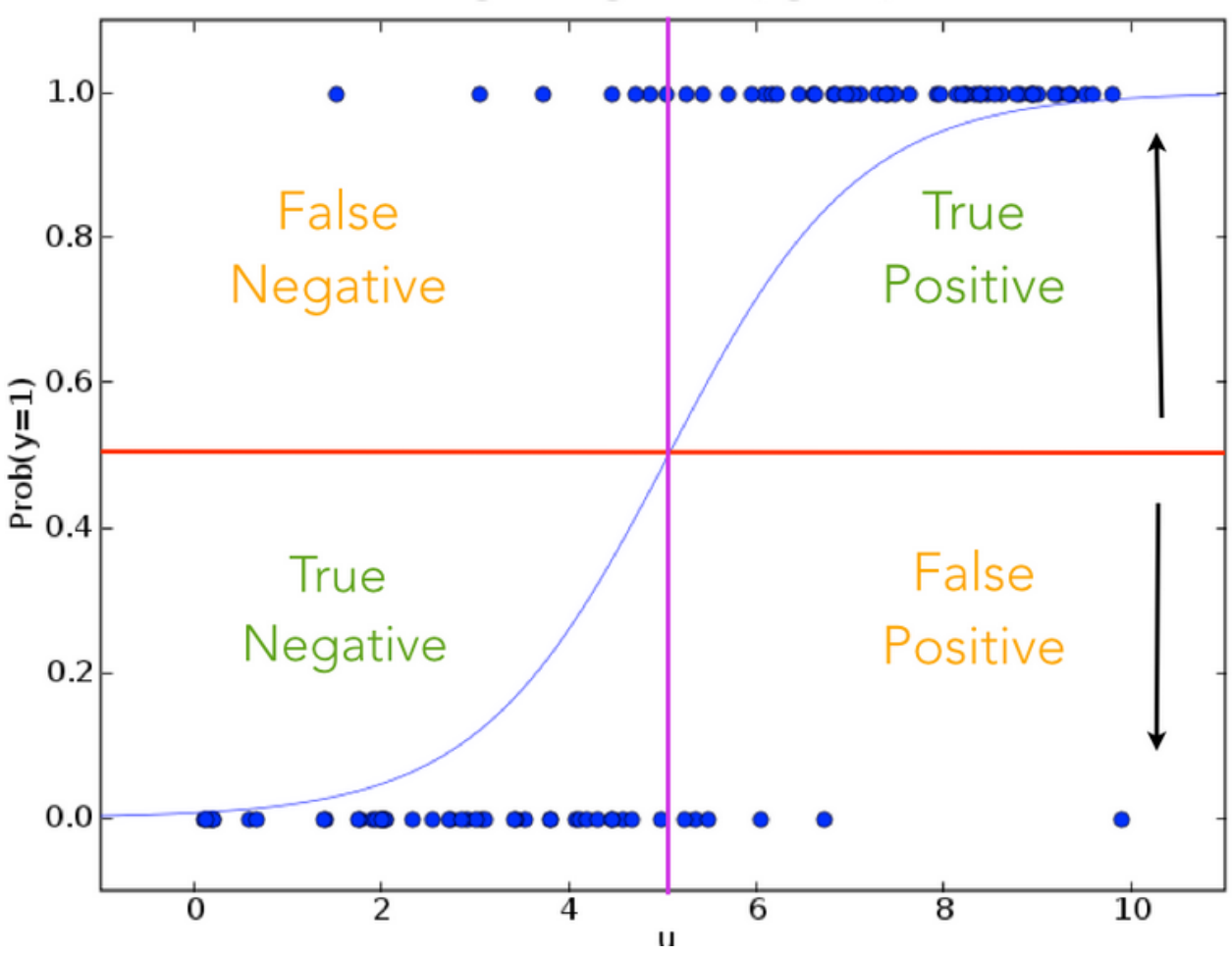
Model score is the probability of class membership ($Y = 1$)

1. Allow T to be the maximum score
2. $TP = 0, FP = 0$
3. For each observation, i:
   - If $\hat{\pi}_i > T \longrightarrow$ increment TP
   - Else $\longrightarrow$ increment FP
4. Add point (FP/N, TP/P) to the ROC Graph

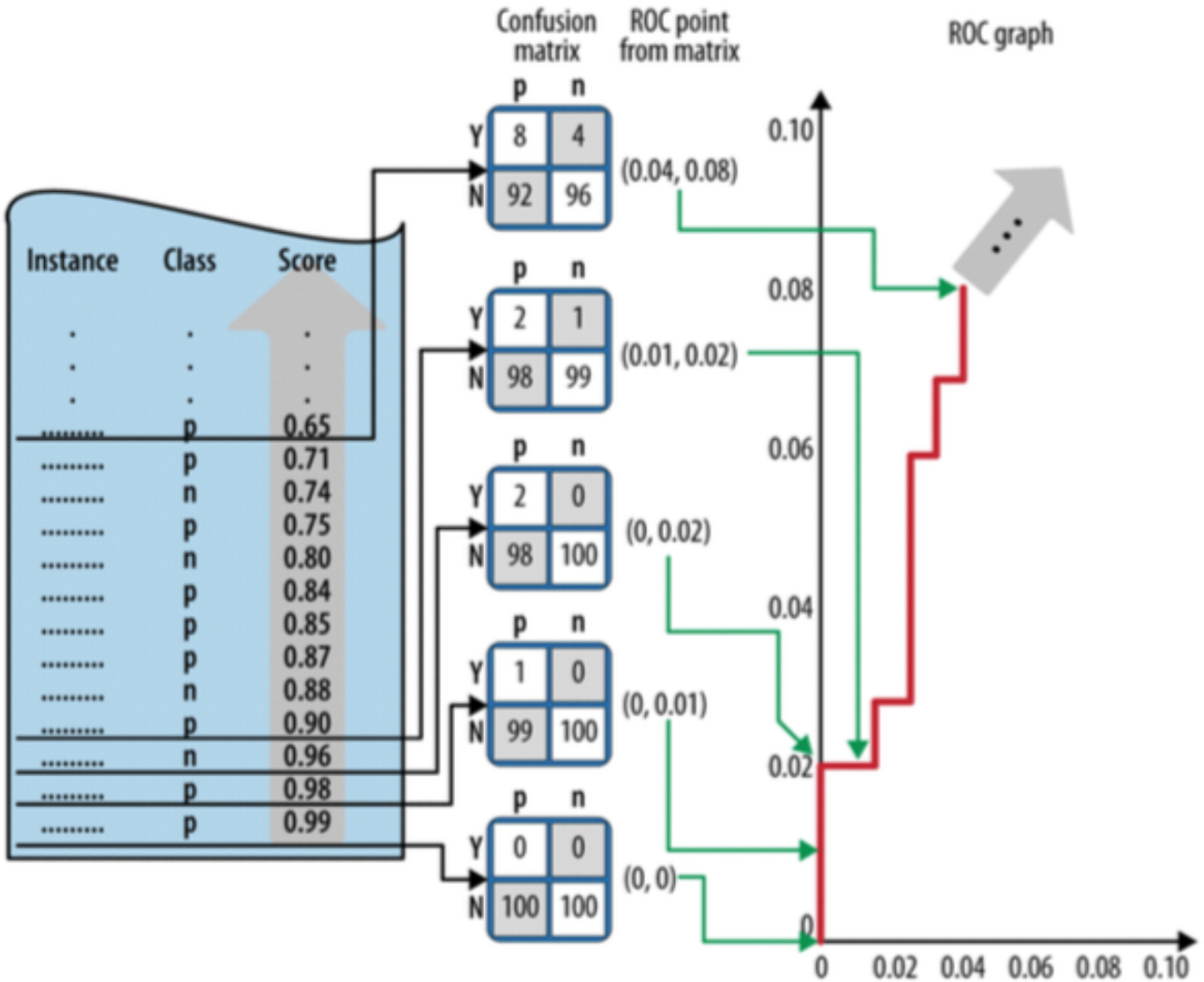
Increment T from max-score to min-score, repeating steps 1-4

Example: Logistic Regression

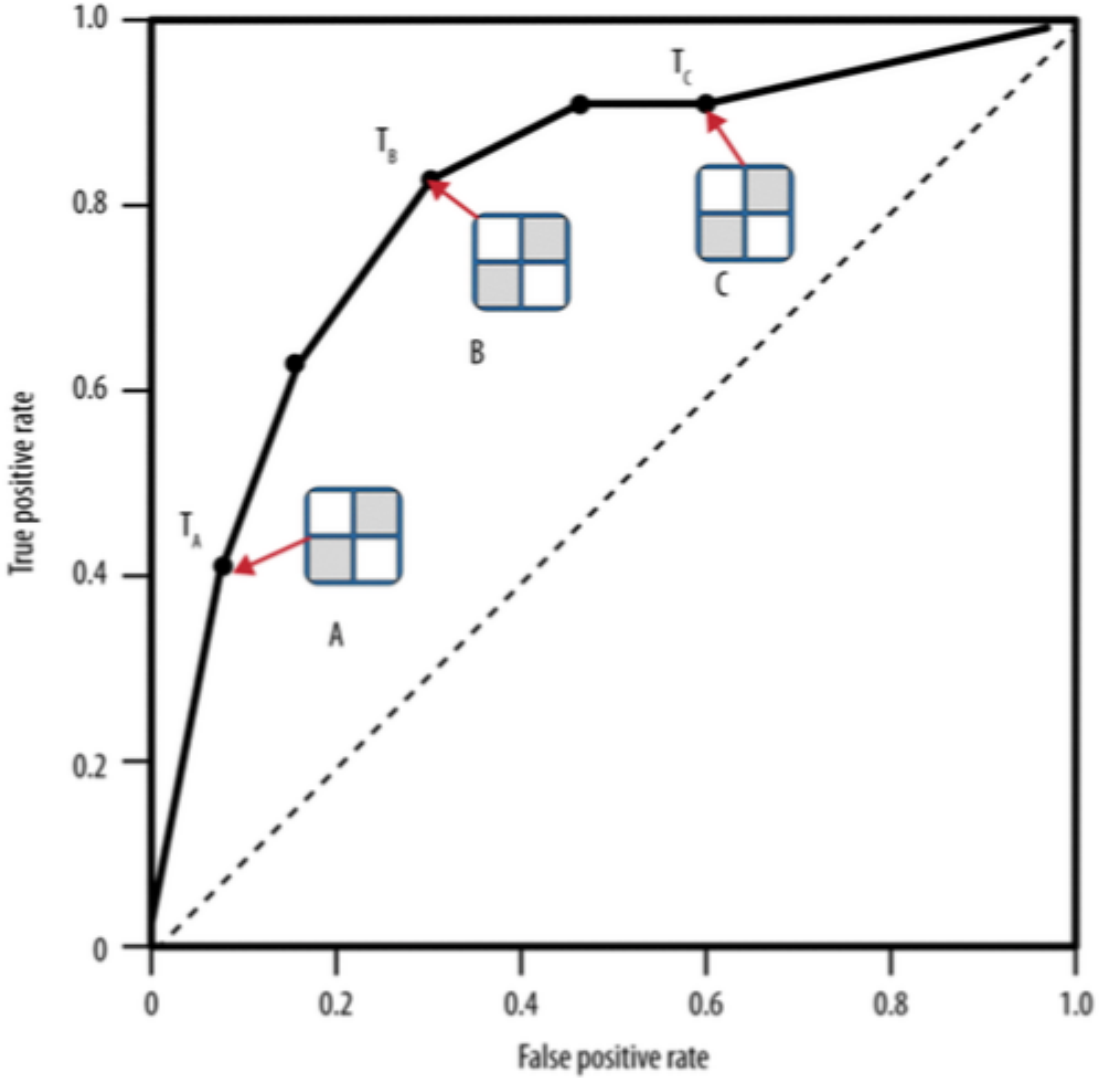
Think of sliding the purple/red line along the sigmoid function



Building the ROC Curve



Sample ROC Curve

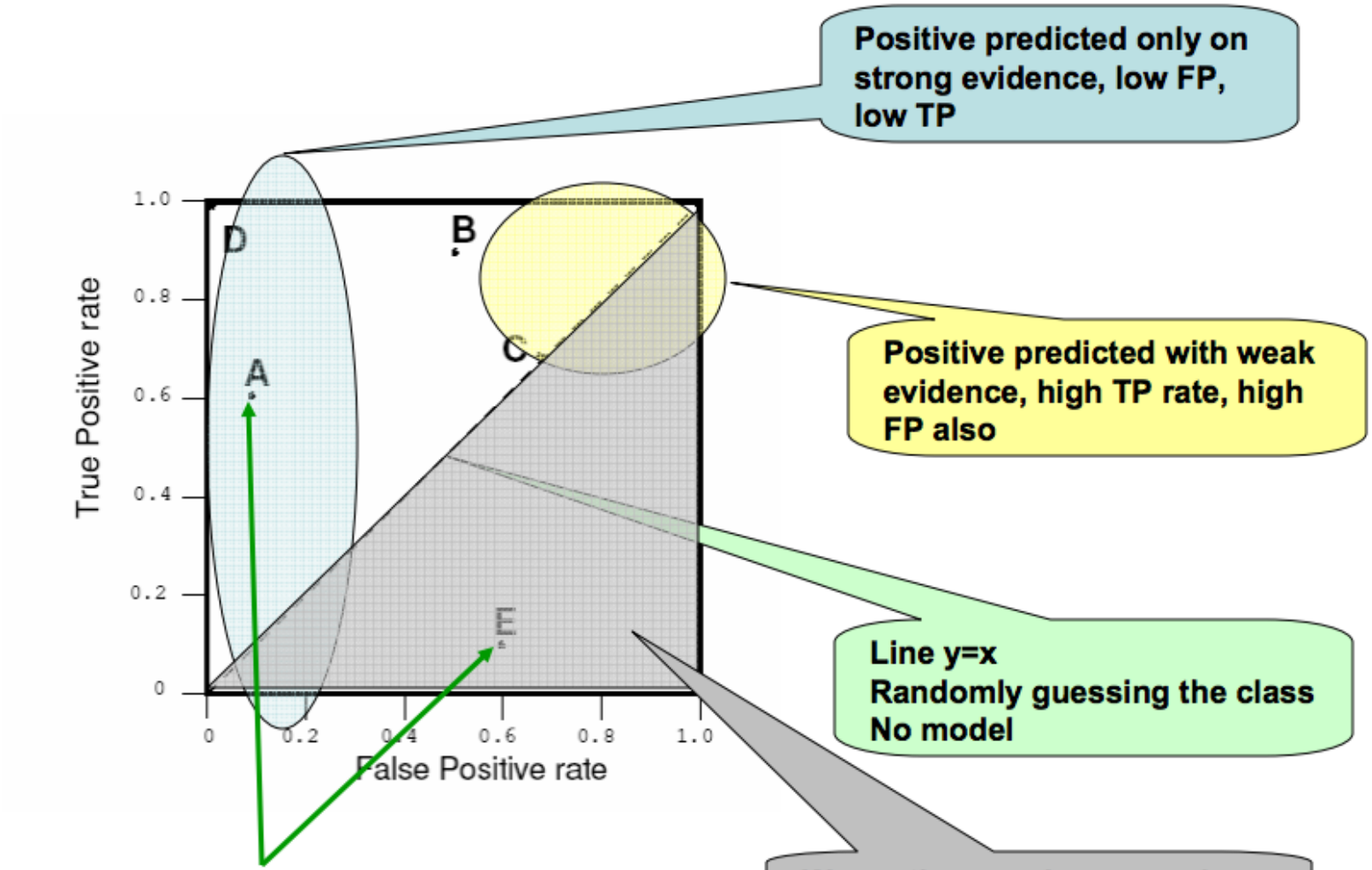


Choosing Between Models

How do we go about choosing a classification model based on the ROC curve?

- Depends on the goal of the model
  - Screening Test
  - Diagnostic Test
- We can examine the regions of the ROC curve based on desired results

Regions of the ROC Curve



ROC Curve for Multiple Classifier Models