Ukraine International Airlines Flight 752

https://en.wikipedia.org/wiki/Ukraine_International_Airlines_Flight_752

Sentiment analysis
Evaluation

Receiver Operating Characteristic
ROC

The ROC curve was first developed by electrical engineers and radar engineers during World War II for detecting enemy objects in battlefield!

Missile attack vs. passenger airplane!

Recall aka True Positive Rate (TPR)

What percentage of positives are captured.

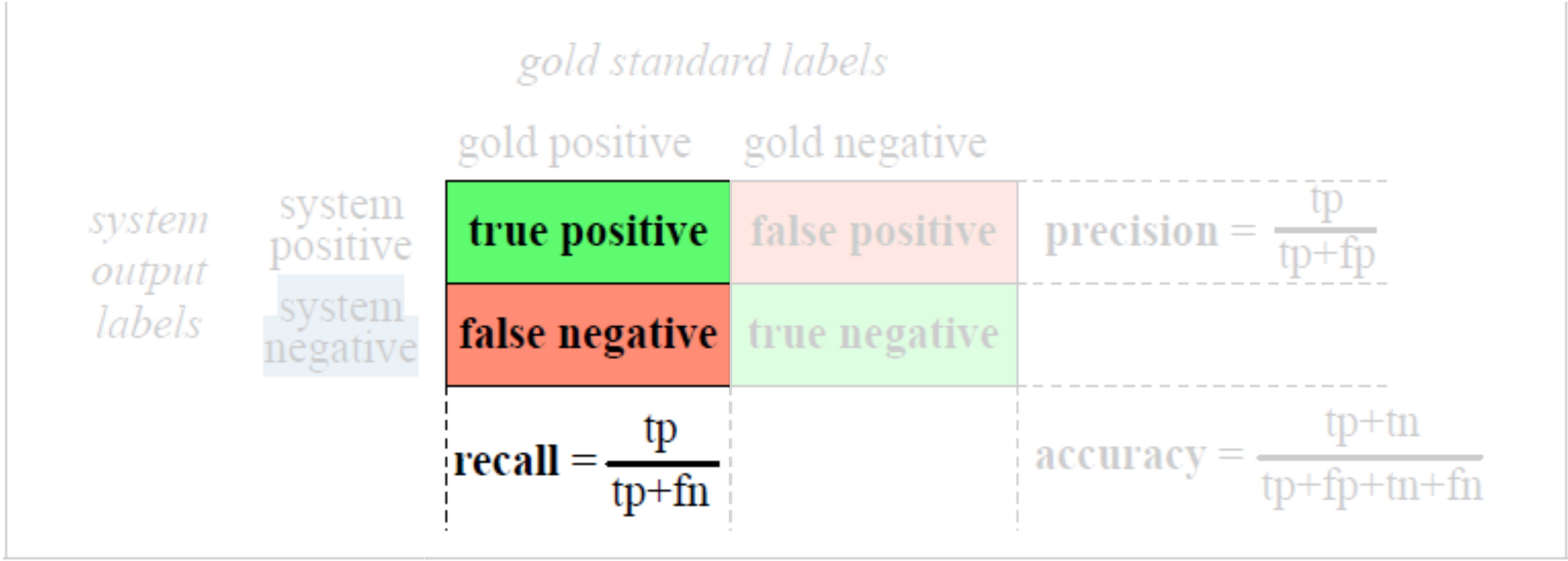


Figure 4.4 Contingency table

False Positive Rate (FPR)

What percentage are *incorrectly* captured as positives!

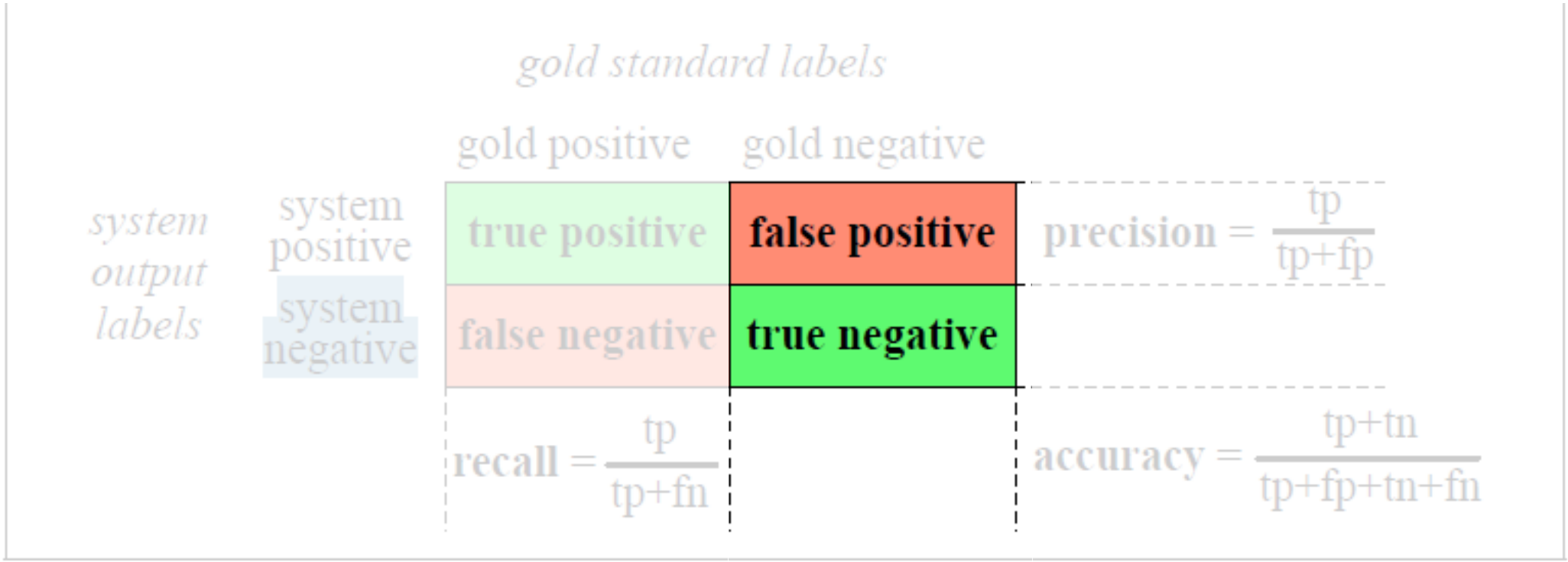


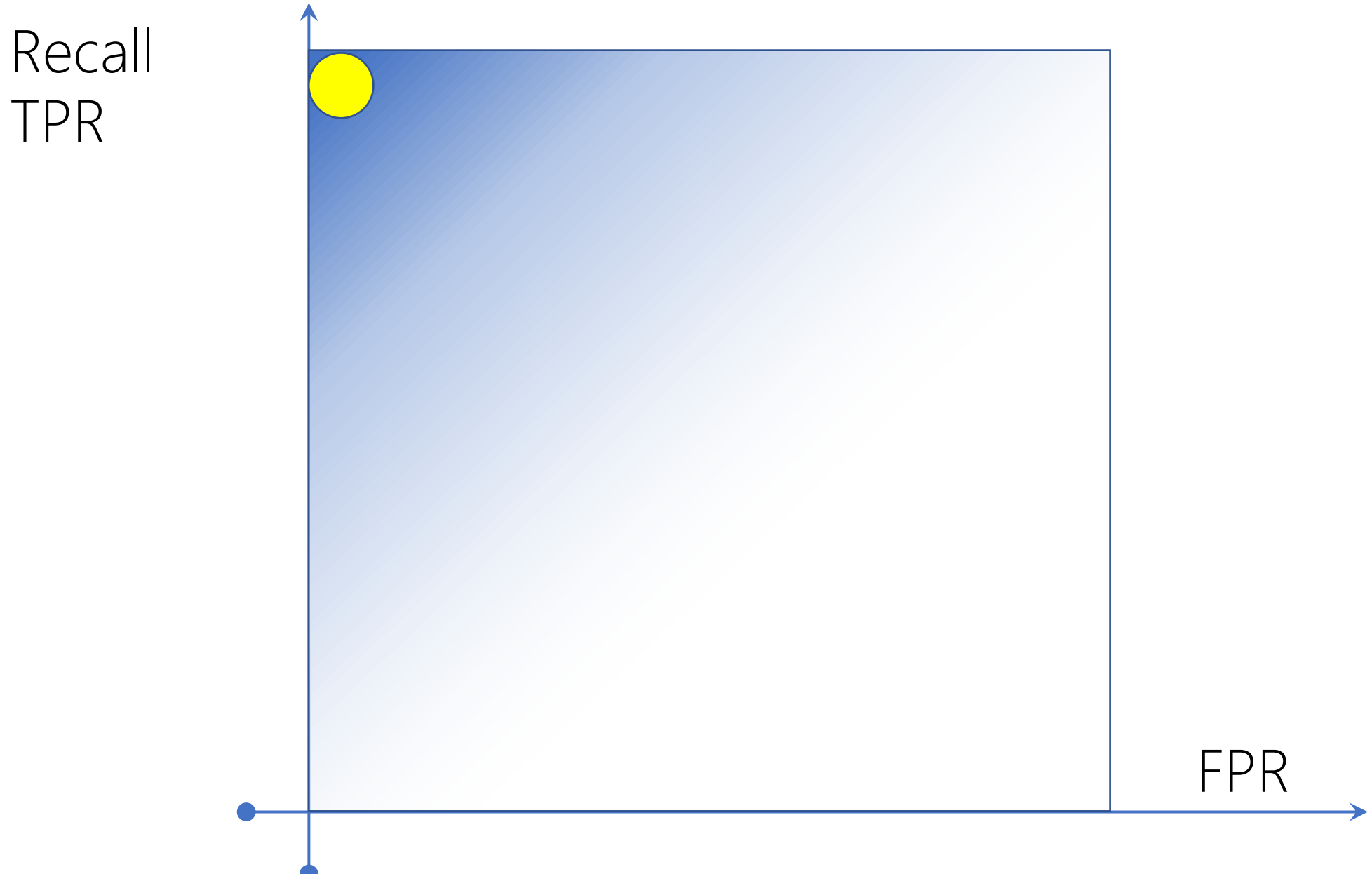
Figure 4.4 Contingency table

Perfect Classifier

	Gold Positive	Gold Negative
Model Positive	N+	0
Model Negative	0	N-

$$\text{TPR}=\frac{N+}{(N+)+0}=1.0 \qquad \text{FPR}=\frac{0}{0+(N-)}=0.0$$

TPR-FPR Curve



Perfect Classifier

	Gold Positive	Gold Negative
Model Positive	N+ ↓	↑ 0
Model Negative	0 ↑	↓ N-

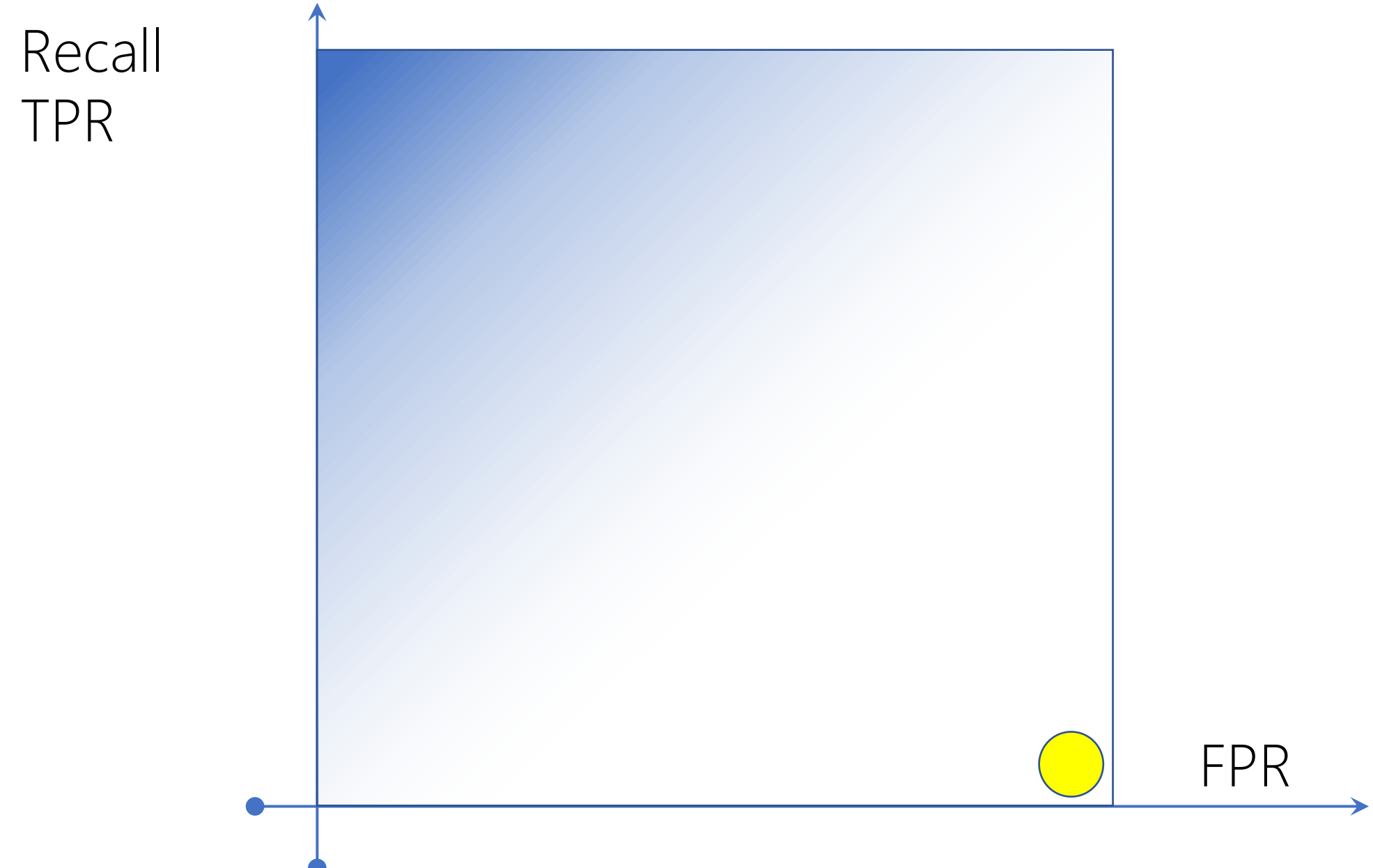
$$\text{TPR}=\frac{N+}{(N+)+0}=1.0 \downarrow \qquad \text{FPR}=\frac{0}{0+(N-)}=0.0 \uparrow$$

Worst Classifier

	Gold Positive	Gold Negative
Model Positive	0	N-
Model Negative	N+	0

$$\text{TPR} = \frac{\mathbf{0}}{(\boldsymbol{N+})+\mathbf{0}} = 0.0 \downarrow \qquad \text{FPR} = \frac{\boldsymbol{N-}}{\mathbf{0}+(\boldsymbol{N-})} = 1.0 \uparrow$$

TPR-FPR Curve

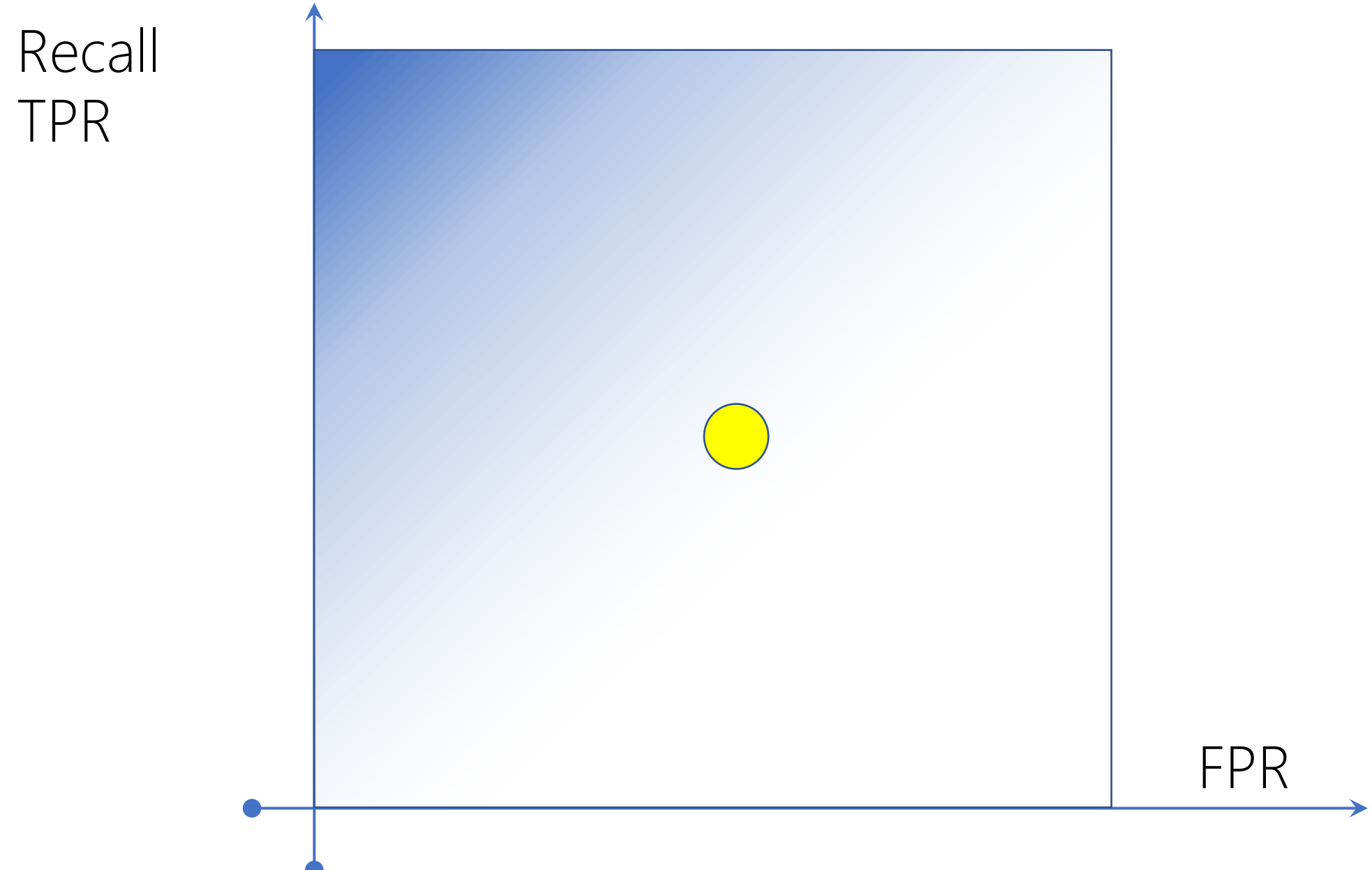


Uniformly Random Classifier

	Gold Positive	Gold Negative
Model Positive	?	?
Model Negative	?	?

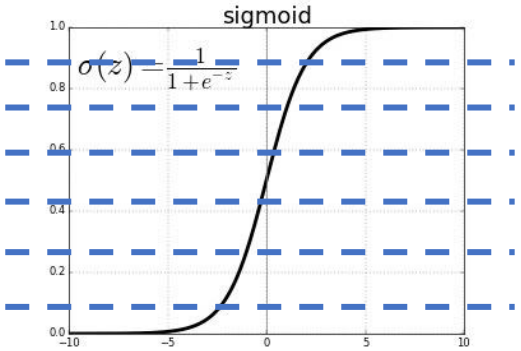
TPR=0.5 FPR=0.5

TPR-FPR Curve



TPR-FPR Curve

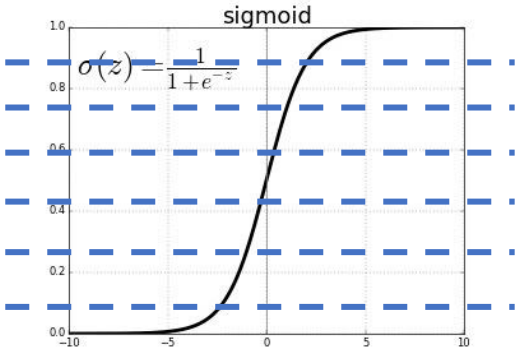
Recall
TPR



$\delta: 0.0 \rightarrow 1.0$

TPR-FPR Curve

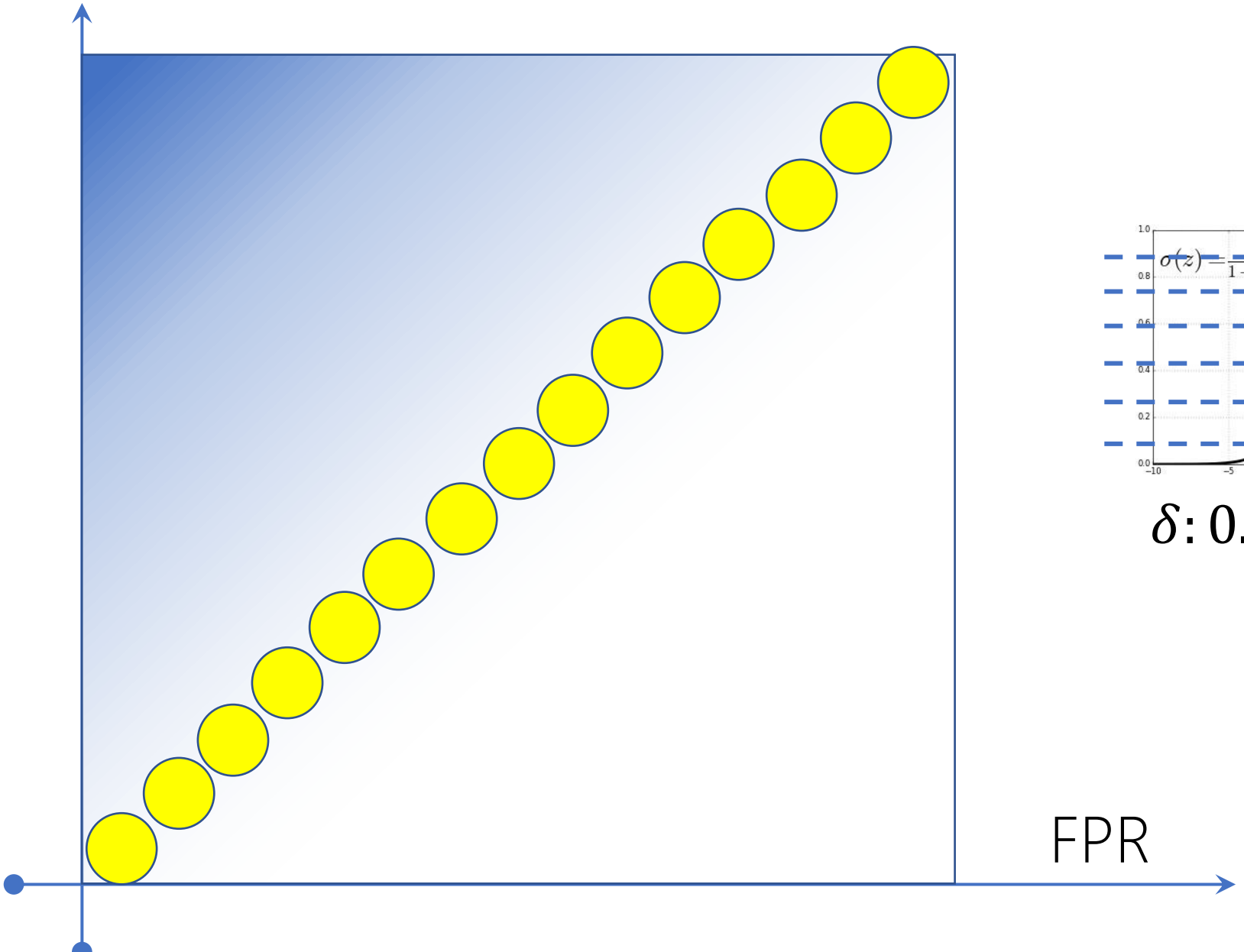
Recall
TPR



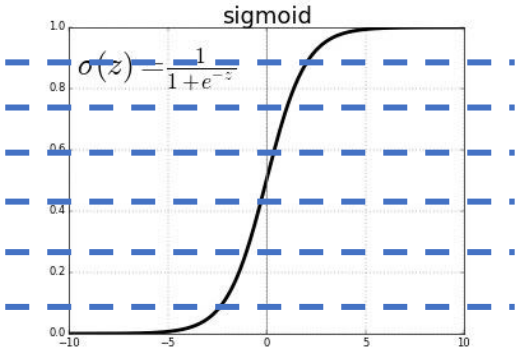
$\delta: 0.0 \rightarrow 1.0$

What is this model?

Recall
TPR



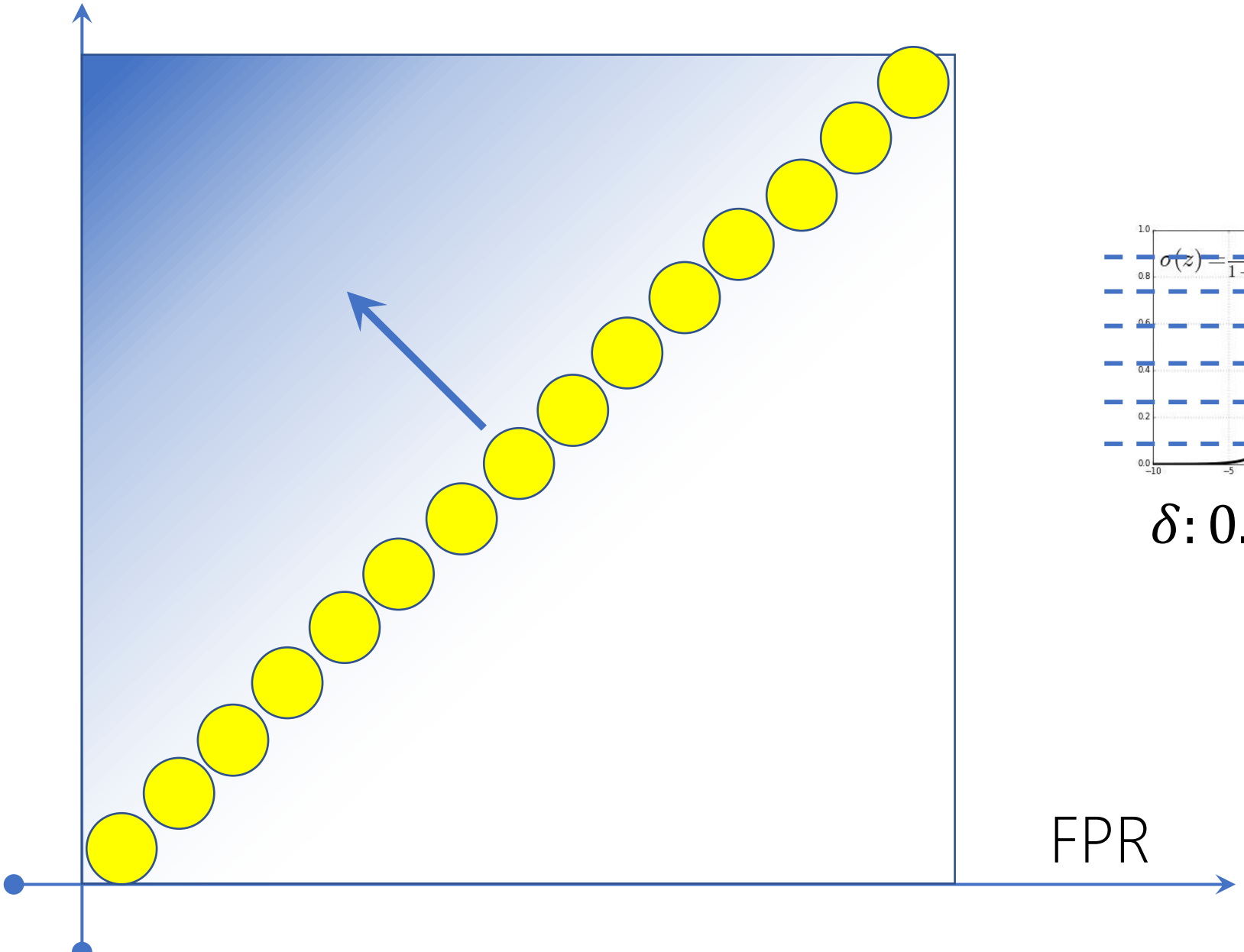
FPR



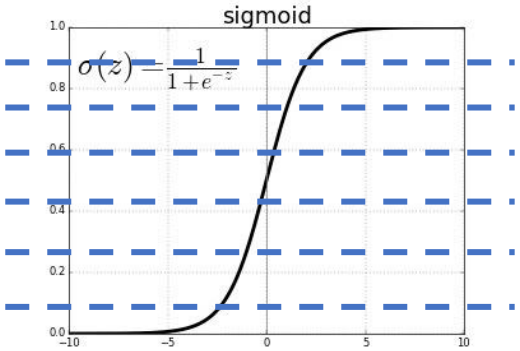
$\delta: 0.0 \rightarrow 1.0$

ROC

Recall
TPR



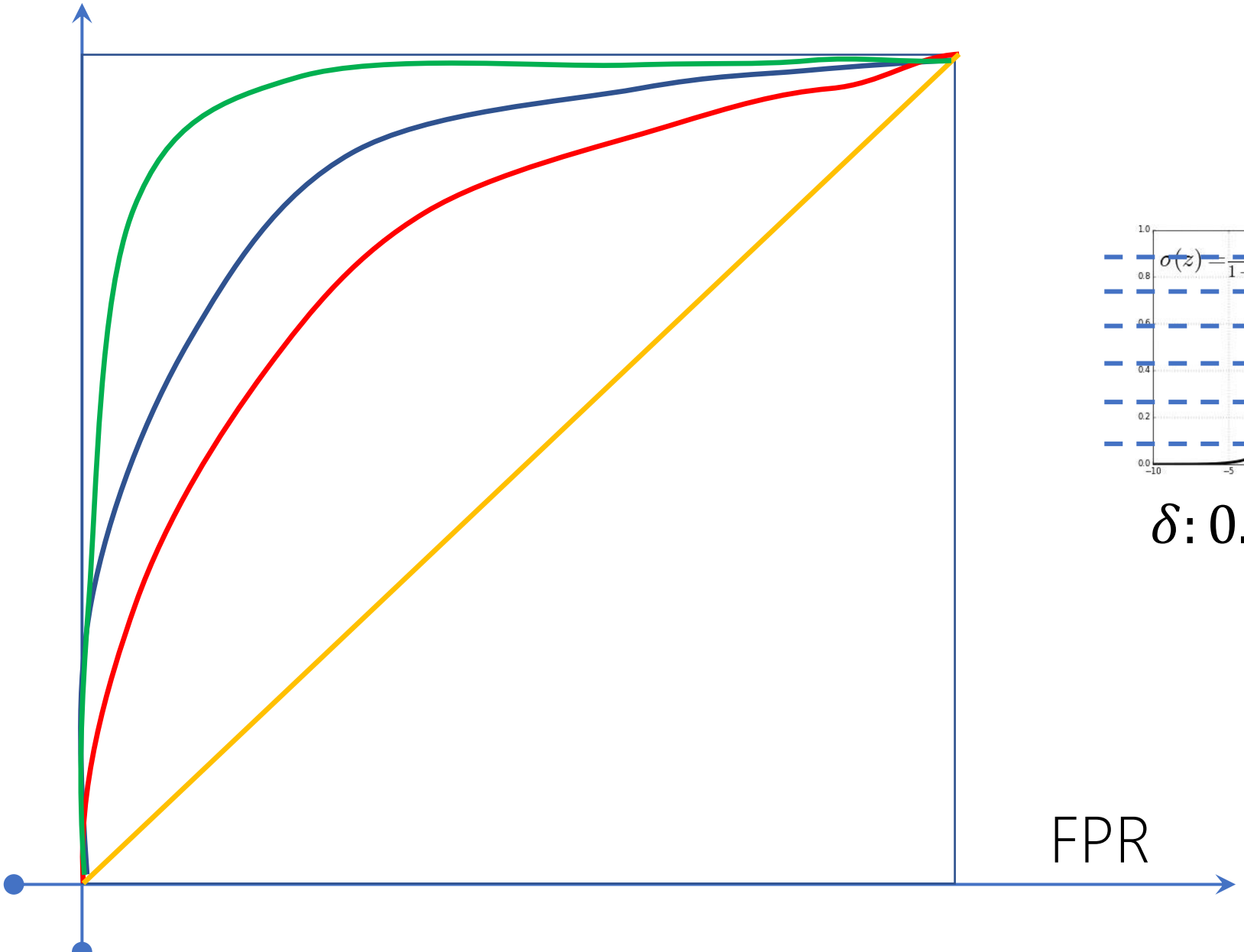
FPR



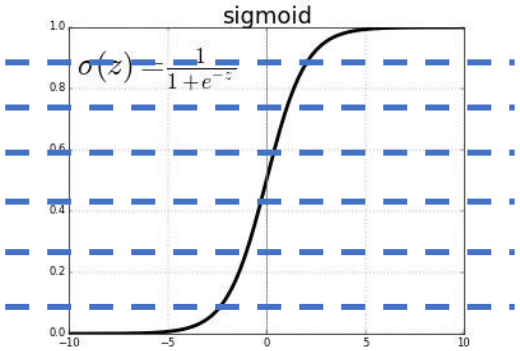
$\delta: 0.0 \rightarrow 1.0$

ROC

Recall
TPR



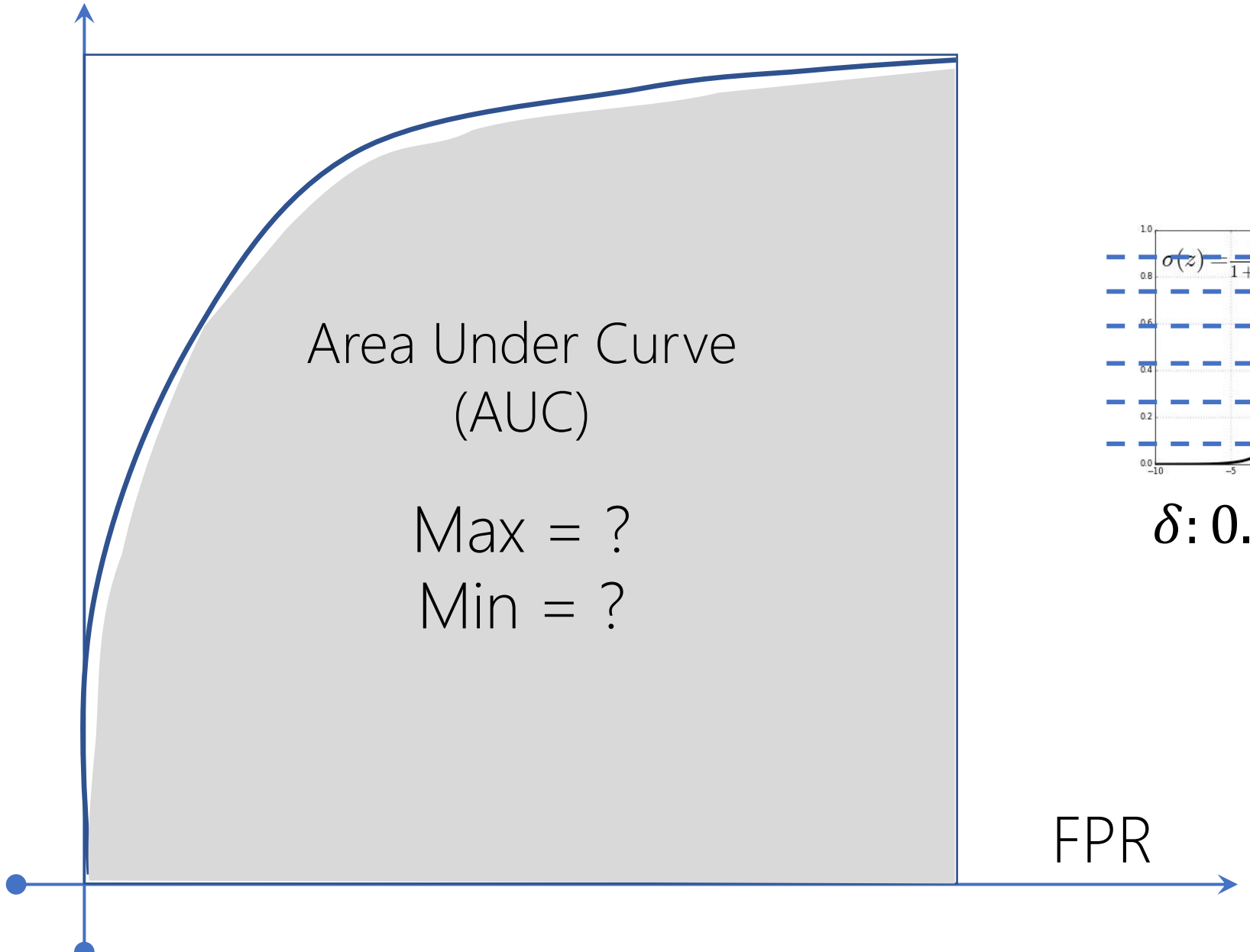
FPR

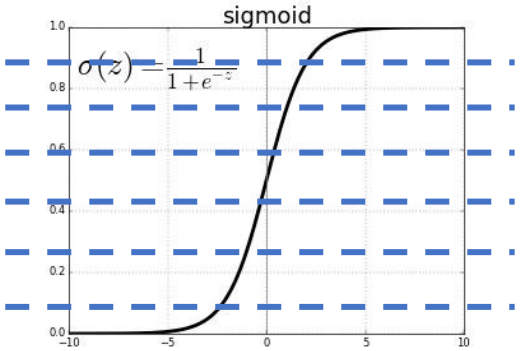


$\delta: 0.0 \rightarrow 1.0$

Area Under Curve (AUC): Single Real Point

Recall
TPR

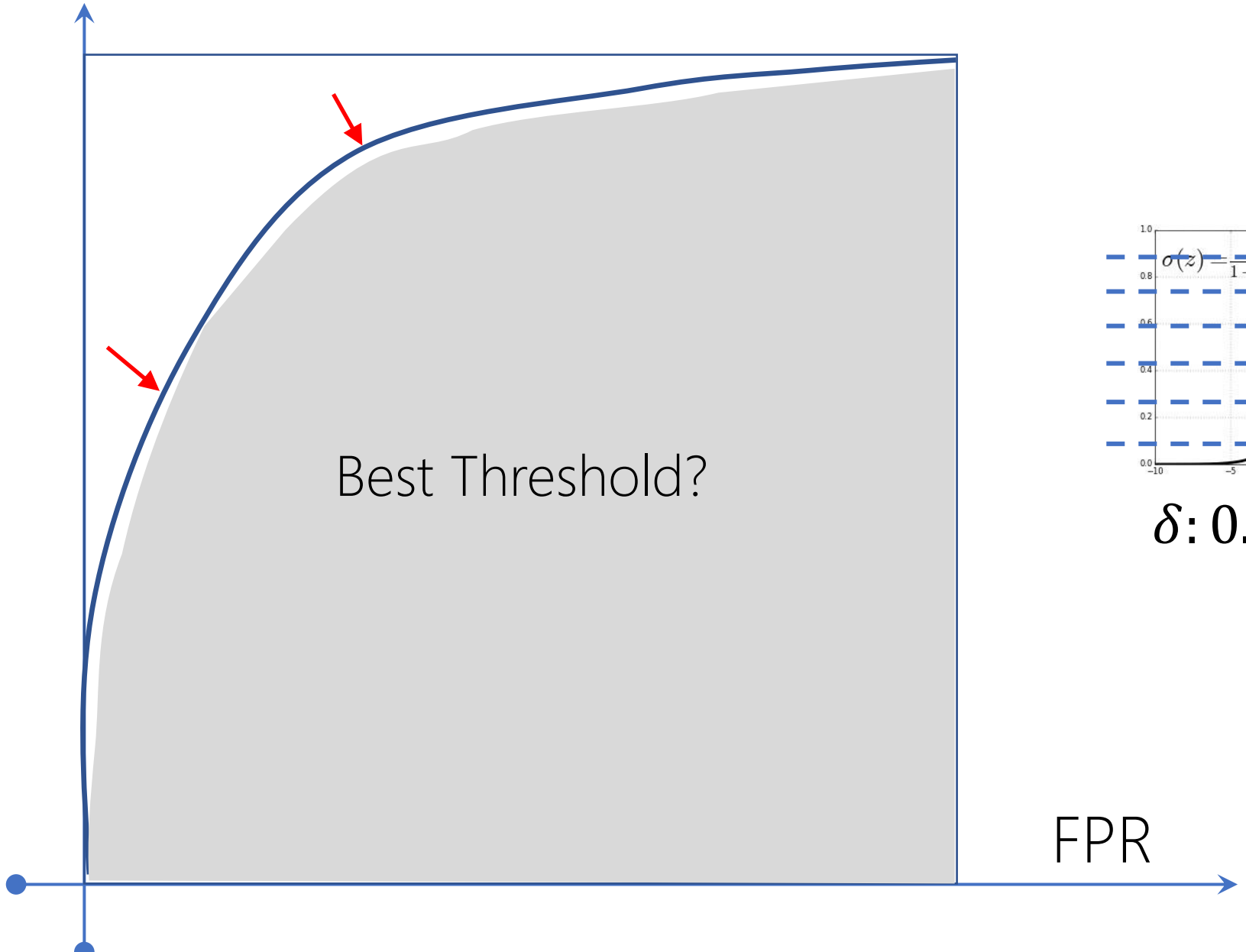


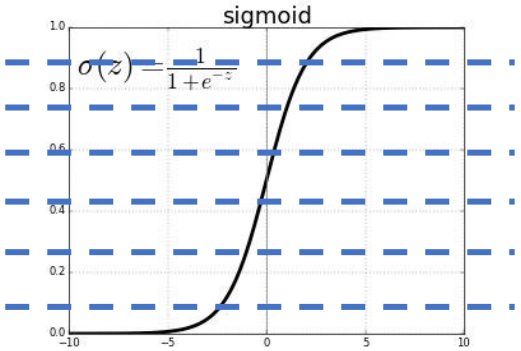


$\delta: 0.0 \rightarrow 1.0$

Area Under Curve (AUC): Single Real Point

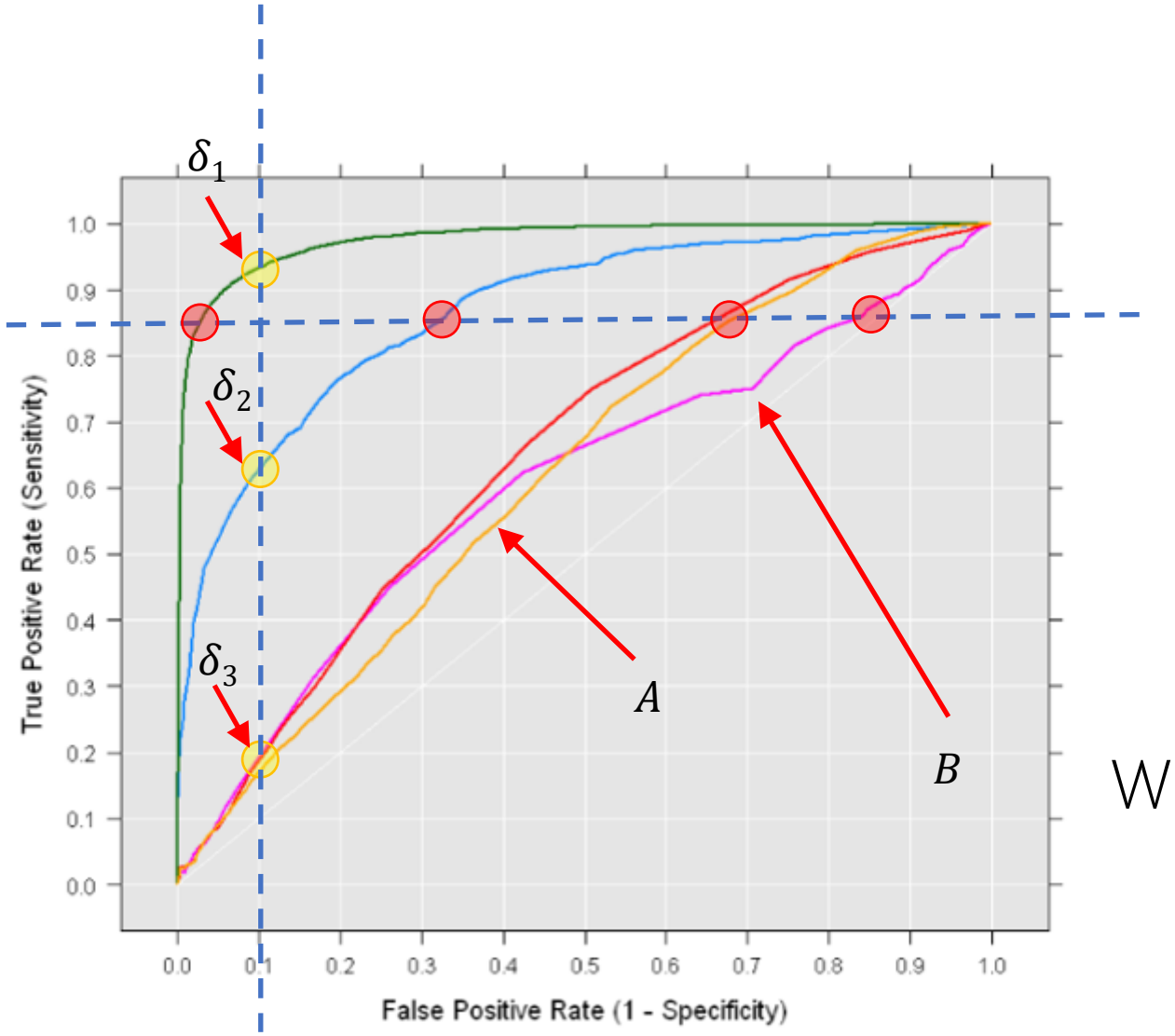
Recall
TPR





$\delta: 0.0 \rightarrow 1.0$

ROC: Model Comparison



Which one? A or B?

Other Metrics

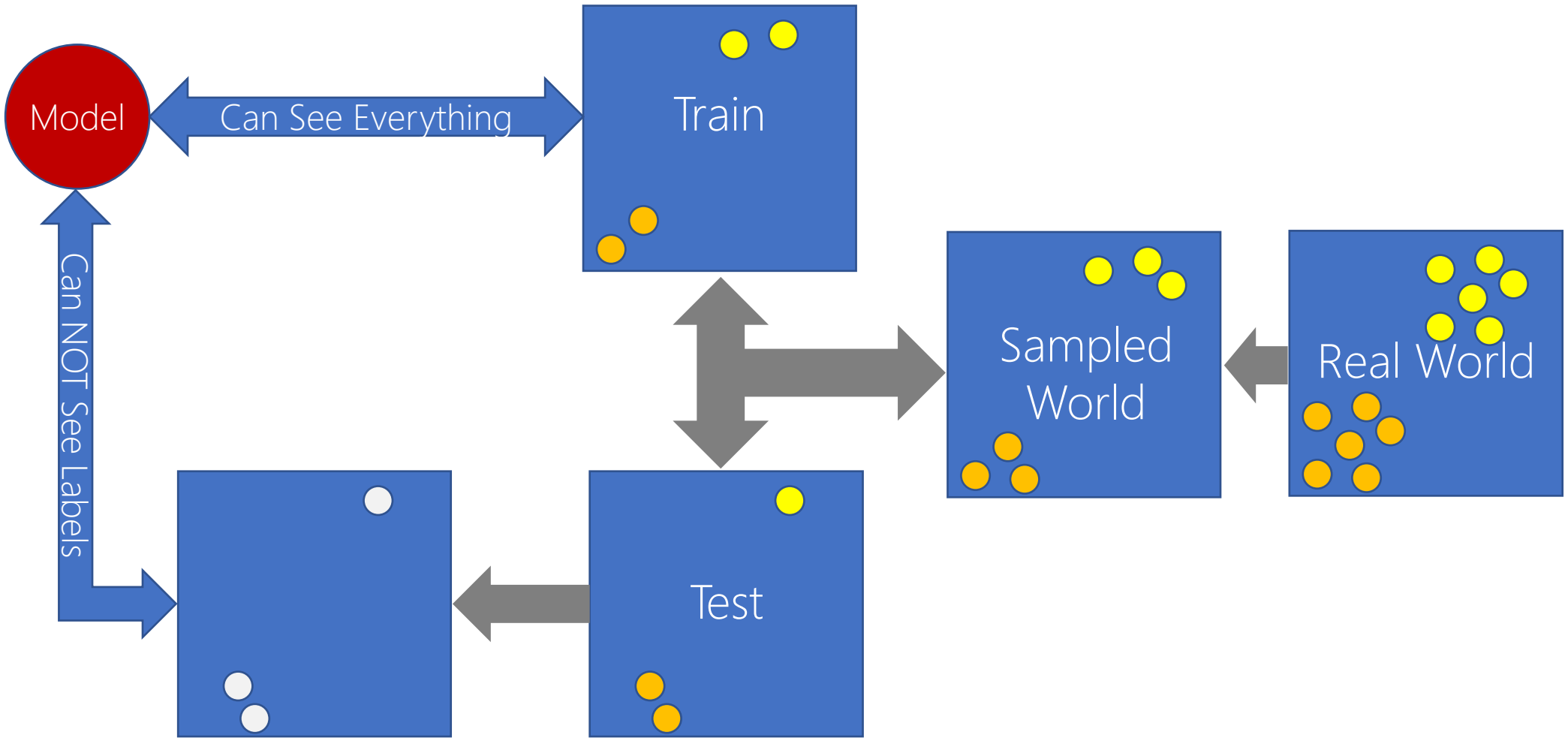
		True condition			
	Total population	Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR−}}$
		False negative rate (FNR), Miss rate = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) = $\frac{\text{FNR}}{\text{TNR}}$	F_1 score = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$

https://en.wikipedia.org/wiki/Precision_and_recall

Calculate the metrics/curves

Labeled Data = {Train} U {Test}

Labeled Data = {Train} U {Test}



Labeled Data = {Train} U {Test}

$$\{\text{Train}\} \cap \{\text{Test}\} \stackrel{?}{=} \emptyset$$

Imbalance Labeled Data = {Train} U {Test}

$$\{\text{Train}\} \cap \{\text{Test}\} \stackrel{?}{=} \emptyset$$

Train and test sets presumably follow same distribution!

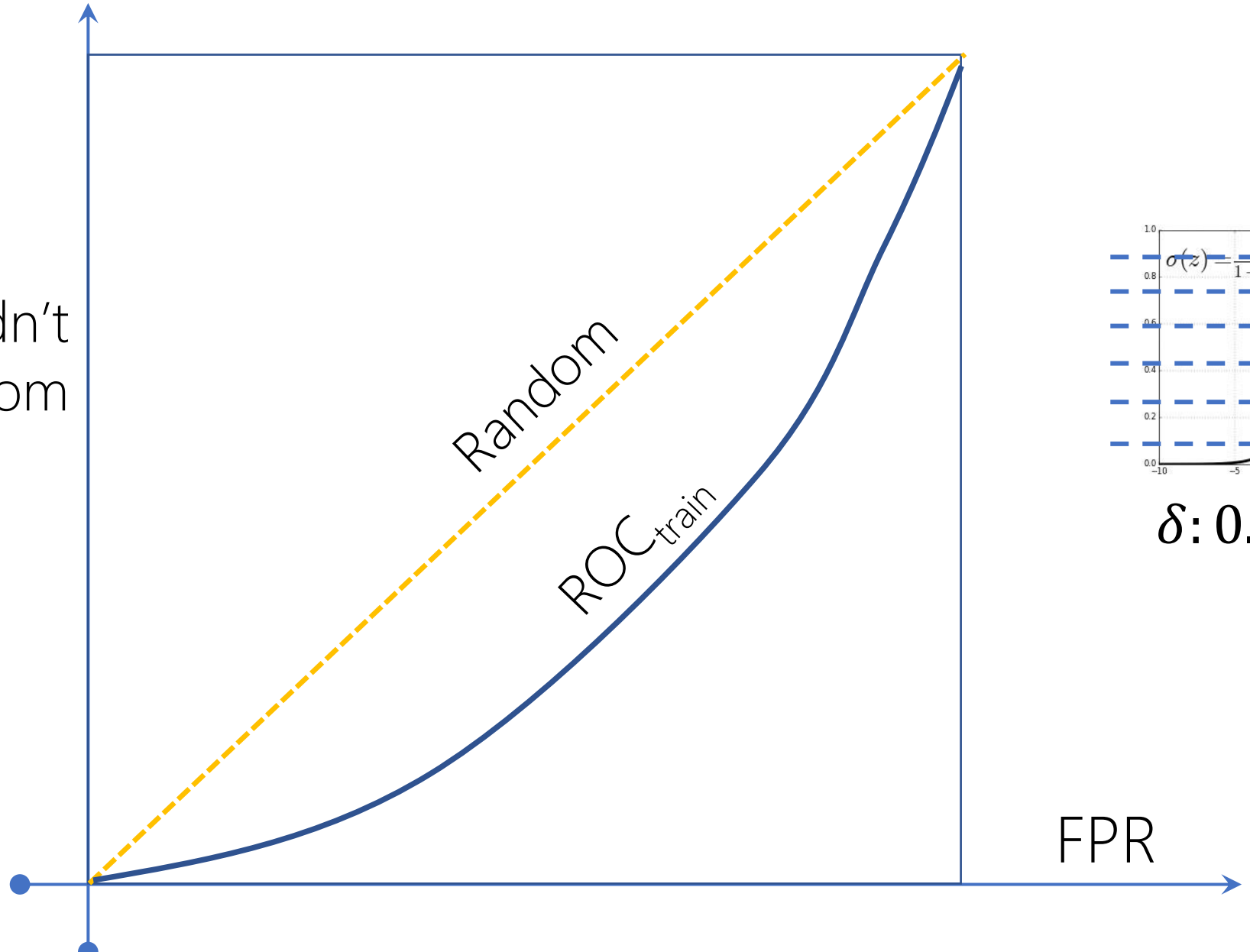
Underfit → Balance fit ← Overfit

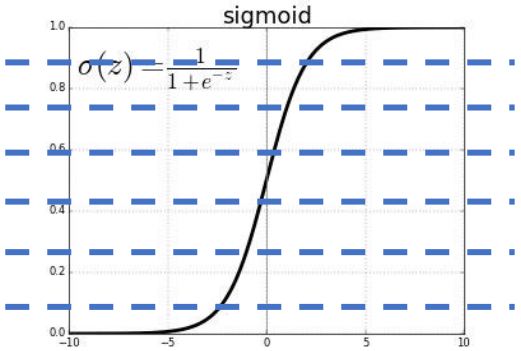
ROC_{train}

Recall
TPR

The model couldn't learn anything from training set!

ROC_{Test} =?





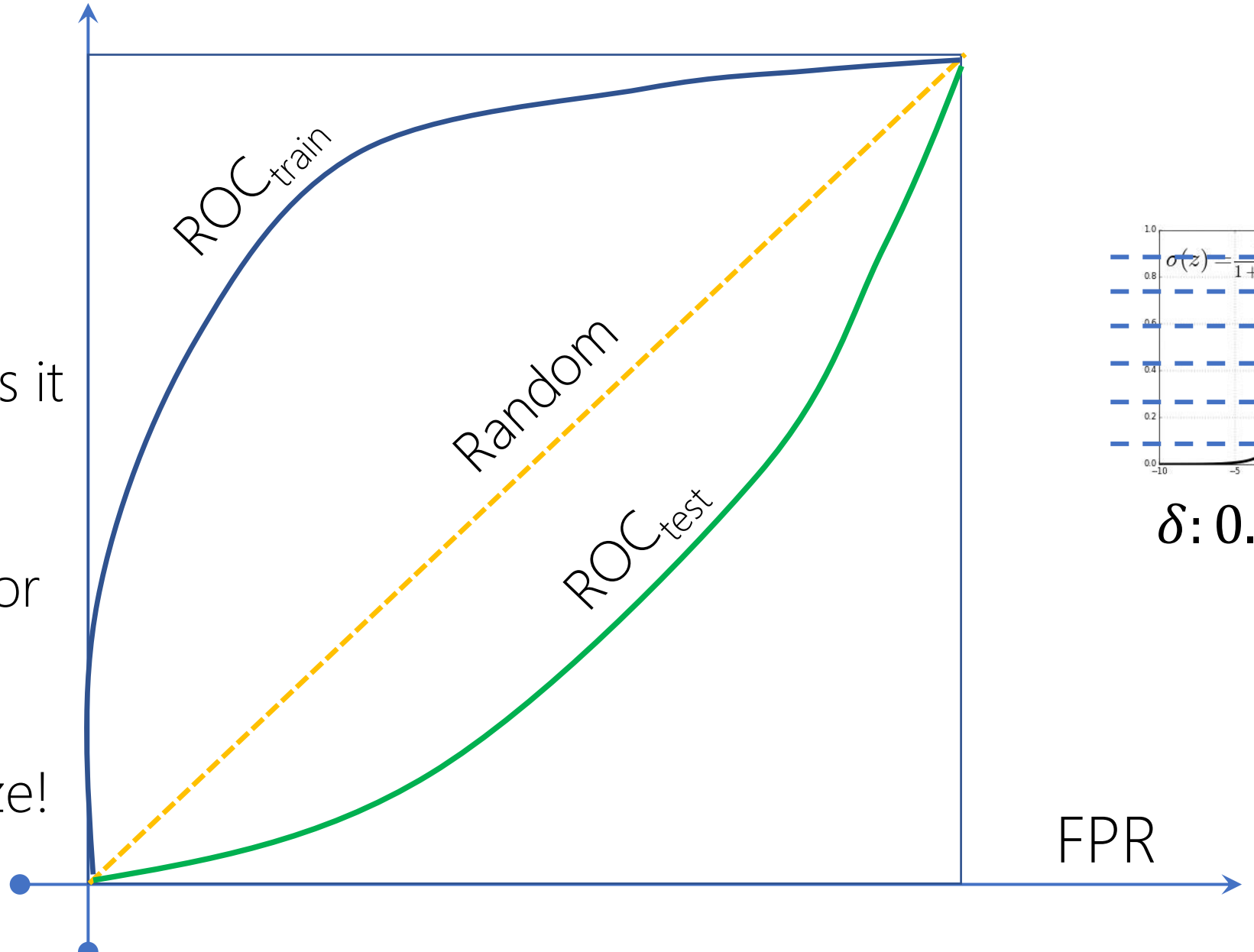
$\delta: 0.0 \rightarrow 1.0$

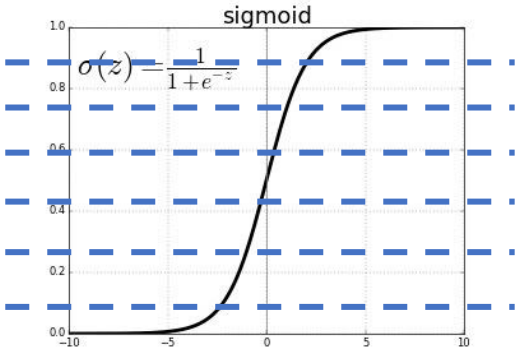
ROC_{test}

Recall
TPR

The model performs well on train set. It means it learnt!

But performs poor in test set. It only know train set. Cannot generalize!





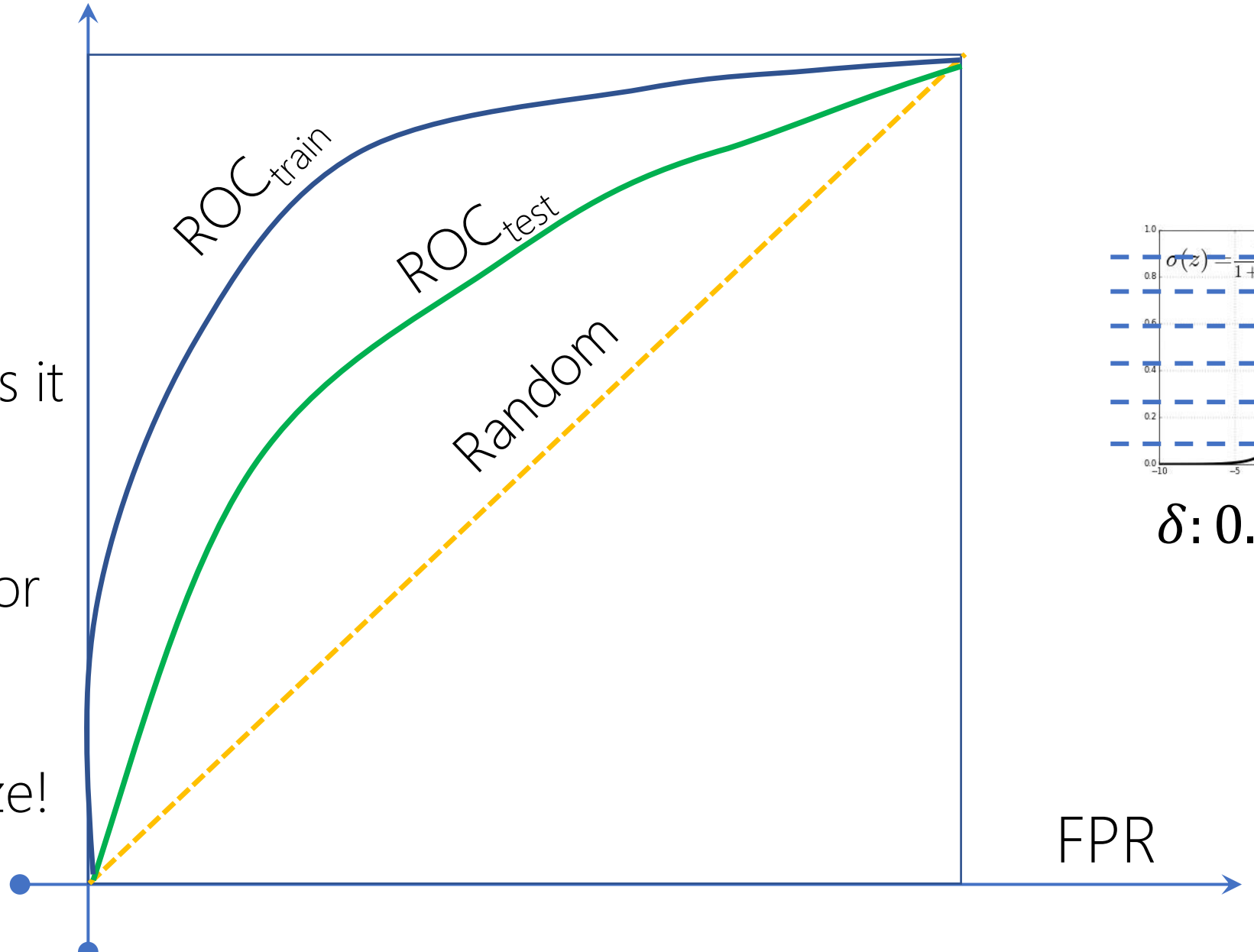
$\delta: 0.0 \rightarrow 1.0$

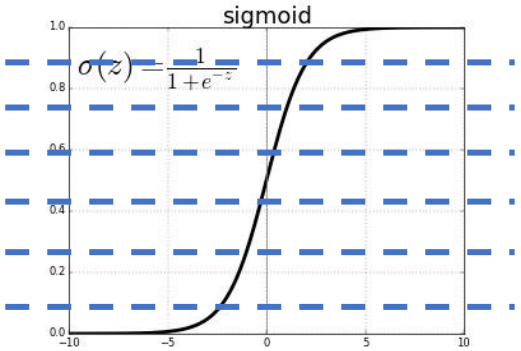
ROC_{test}

Recall
TPR

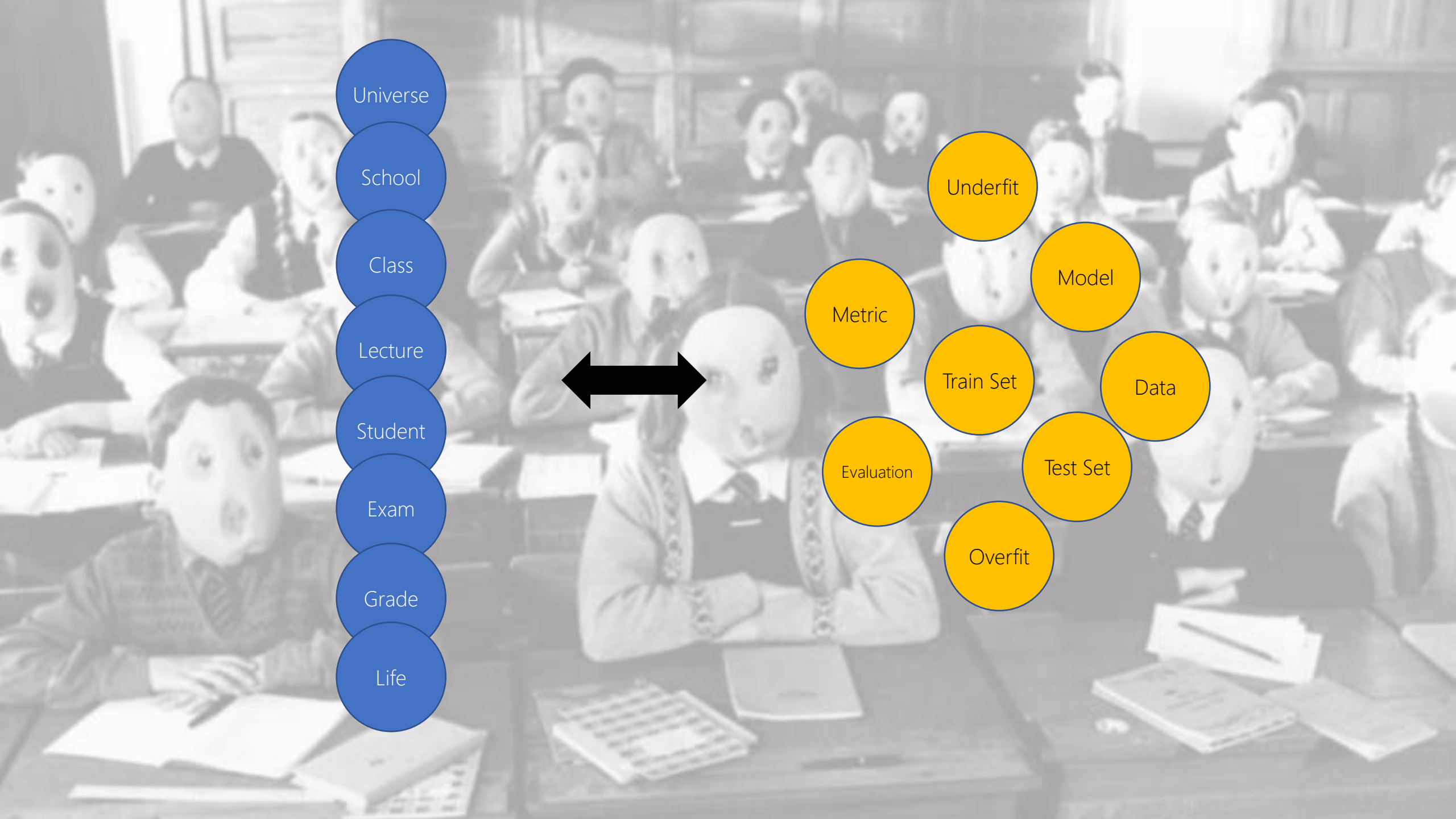
The model performs well on train set. It means it learnt!

But performs poor in test set. It only know train set. Cannot generalize!





$\delta: 0.0 \rightarrow 1.0$


Universe
School
Class
Lecture
Student
Exam
Grade
Life
Underfit
Model
Metric
Train Set
Data
Evaluation
Test Set
Overfit

Model Tuning

Find the best running settings of the mode

- #layers
- Activation functions
- Probs. assumption

Model Tuning

Find the best running settings of the mode

- Checking the performance of model on Train and Test
- For all different possibilities

Blind grid search! Brute-force

Model Tuning

Find the best running settings of the mode

- Learn the performance of model on Train and Test
- For all different possibilities

Guided grid search!

Model Tuning

Find the best running settings of the mode

- Learn the performance of model on Train and Test
- For all different possibilities

Guided grid search!