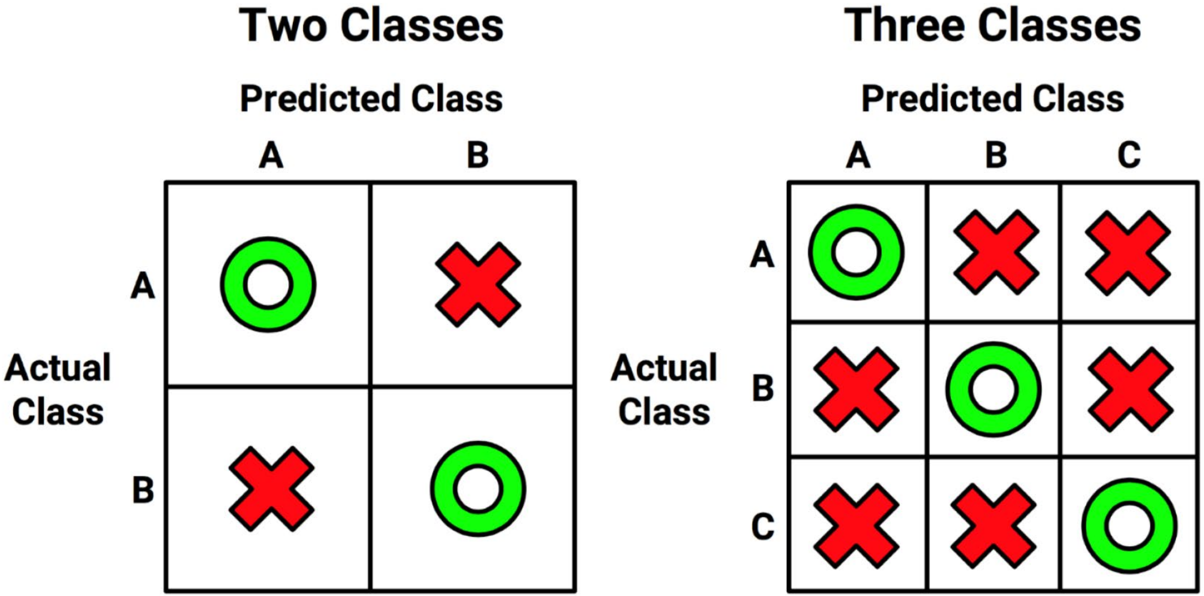
Data Prediction Model and Machine Learning

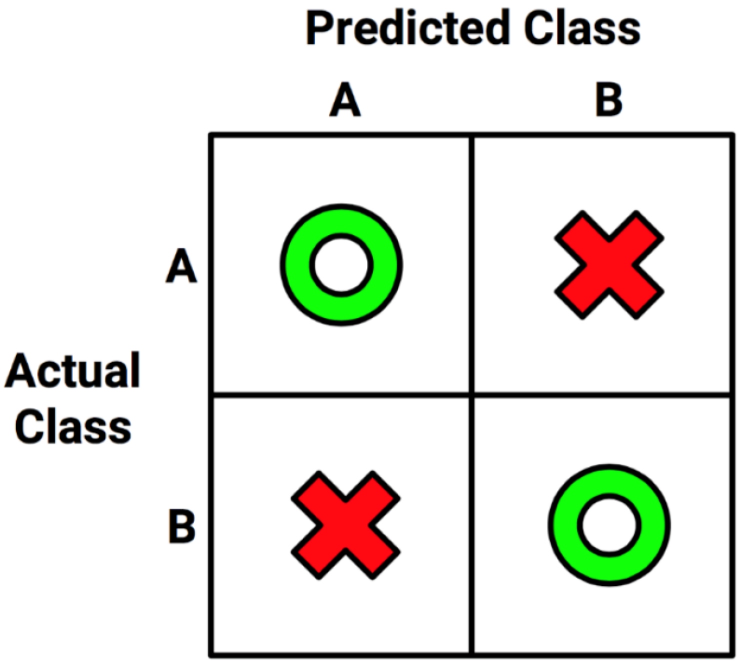
Online course #11

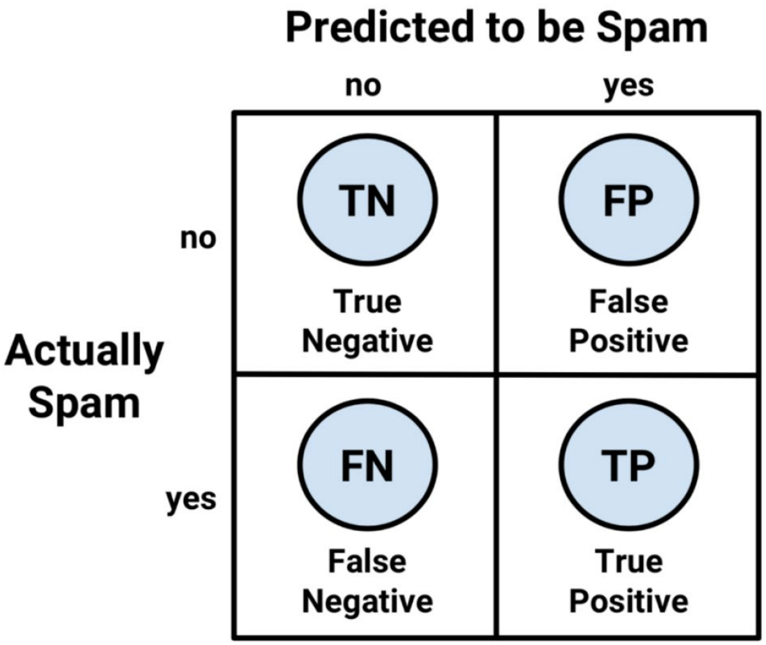
Model improvement

Accuracy: Percentage of right prediction



Learning more about confusion matrix





Learning more about confusion matrix

Ground Truth

Prediction

	Positive (Spam)	Negative (Ham)
Positive (Spam)	TP	FP
Negative (Ham)	FN	TN

- Accuracy (정확도) = (TP + TN) / All
- Error rate (오류율) = (FP + FN) / All

Beyond Accuracy: kappa statistic

- Kappa statistic adjusts accuracy by controlling the likelihood of accidentally making accurate predictions

  e.g.) In severely imbalanced data (like 90% positive), high accuracy can be obtained by just one-sided predictions

Ratio of actual agreement
= **Accuracy**

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Expected agreement between actual and predicted values, assuming that they were chosen at random

Beyond Accuracy: kappa statistic

		Ground Truth	
		Positive (Spam)	Negative (Ham)
Prediction	Positive (Spam)	20	10
	Negative (Ham)	30	150

180/210 * 160/210 = 0.65

Probability of making a prediction as ham while it is actually ham

+

Probability of making a prediction as spam while it is actually spam

30/210 * 50/210 = 0.03

Ratio of actual agreement
= **Accuracy**

170/210 = 0.81

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Expected agreement between actual and predicted values, assuming that they were chosen at random

(0.81 – 0.68) / (1-0.68)
= 0.13 / 0.32
= 0.41

Beyond Accuracy: Sensitivity vs. Specificity

Finding a **useful classifier** requires a balance between overly **conservative** and overly **aggressive** predictions.

i.e.) Spam filter

Trade-off between

99% of Spams are filtered correctly but 5% of Ham is mis-filtered

Vs.

80% of Spams are filtered correctly but only 0.1% of Ham is mis-filtered

Beyond Accuracy: Sensitivity vs. Specificity

Ground Truth

		Positive (Spam)	Negative (Ham)
Prediction	Positive (Spam)	TP	FP
	Negative (Ham)	FN	TN

- **Sensitivity**: Correctly classified positive rate
- TPR (True Positive Rate) = TP / (TP + FN)

- **Specificity**: Correctly classified negative rate
- TNR (True Negative Rate) = TN / (TN + FP)

Beyond Accuracy: Sensitivity vs. Specificity

Trade-off between

99% of Spams are filtered correctly but 5% of Ham is mis-filtered

Vs.

80% of Spams are filtered correctly but only 0.1% of Ham is mis-filtered

Trade-off between

Sensitivity 99% and **Specificity** 95%

Vs.

Sensitivity 80% and **Specificity** 99.5%

Beyond Accuracy: Precision vs. Recall

		Ground Truth	
		Positive (Spam)	Negative (Ham)
Prediction	Positive (Spam)	TP	FP
	Negative (Ham)	FN	TN

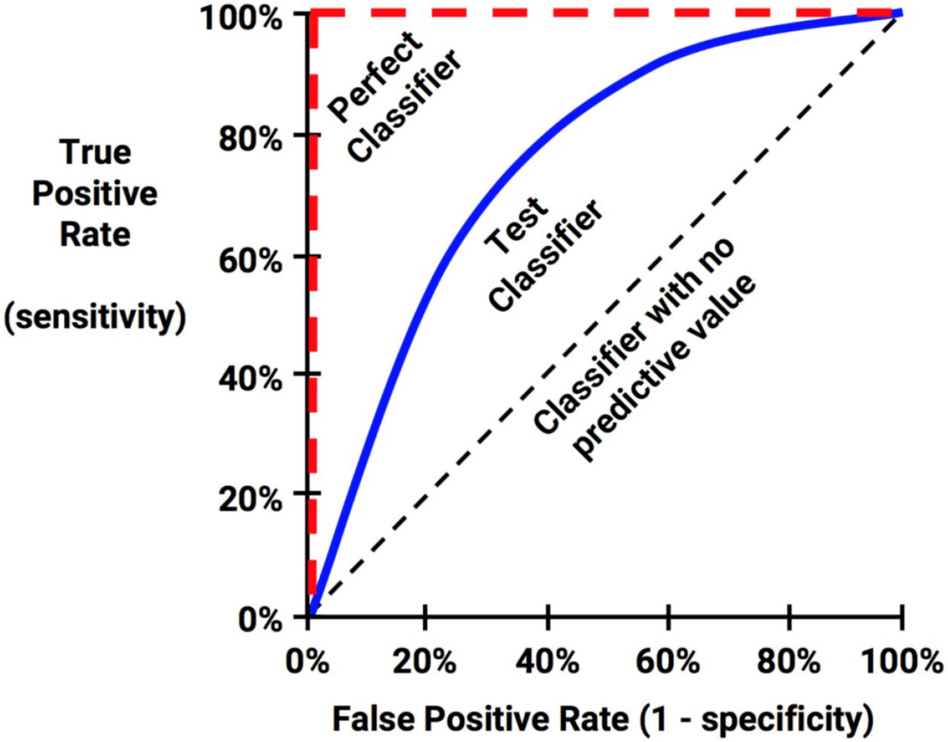
- **Precision**: How accurate is it when predicting positives
  = TP / (TP + FP)

- **Recall**: How perfectly classified positive values
  =TP / (TP + FN)

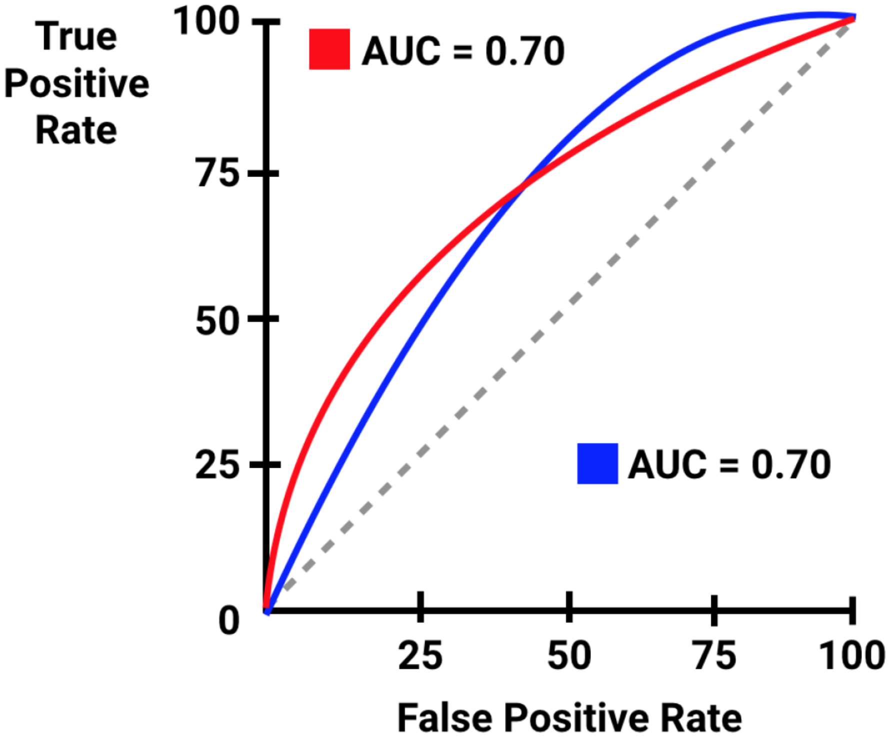
ROC (Receiver Operating Characteristic) **curve:** How to draw?

- **Sensitivity**: Correctly classified positive rate
- TPR (True Positive Rate) = TP / (TP + FN)

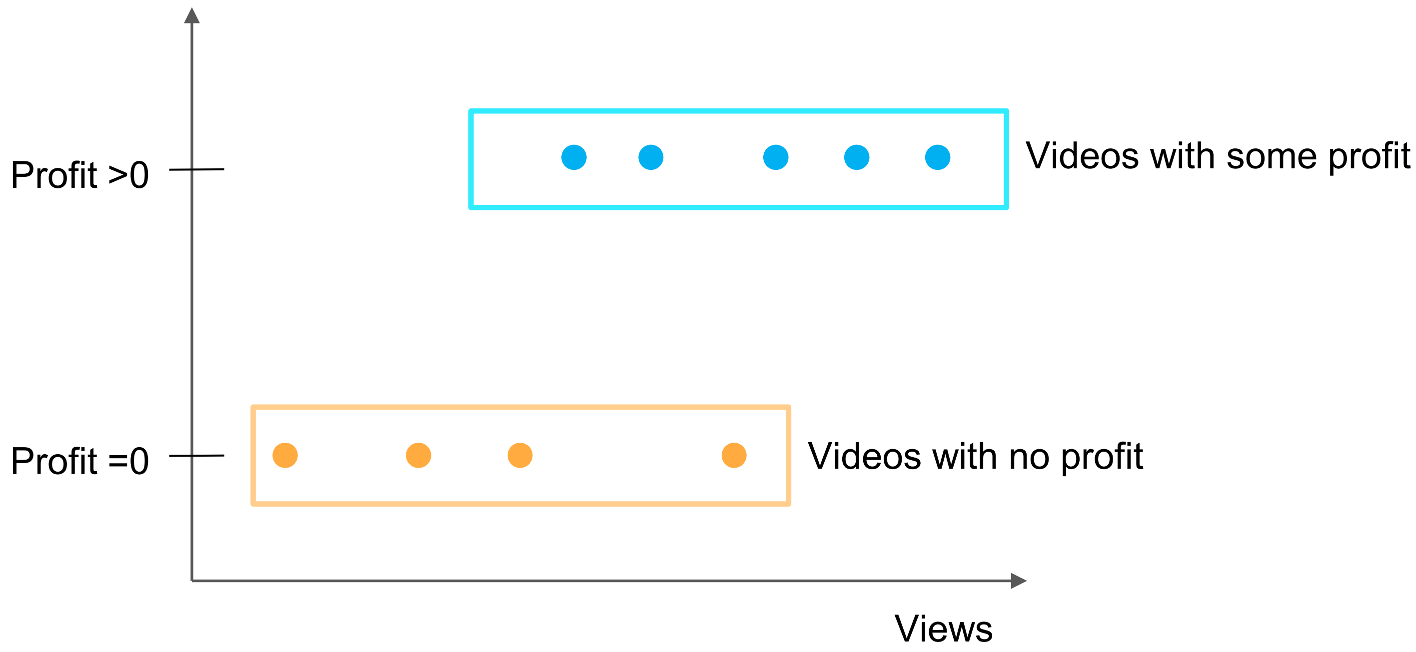
- **Specificity**: Correctly classified negative rate
- TNR (True Negative Rate) = TN / (TN + FP)



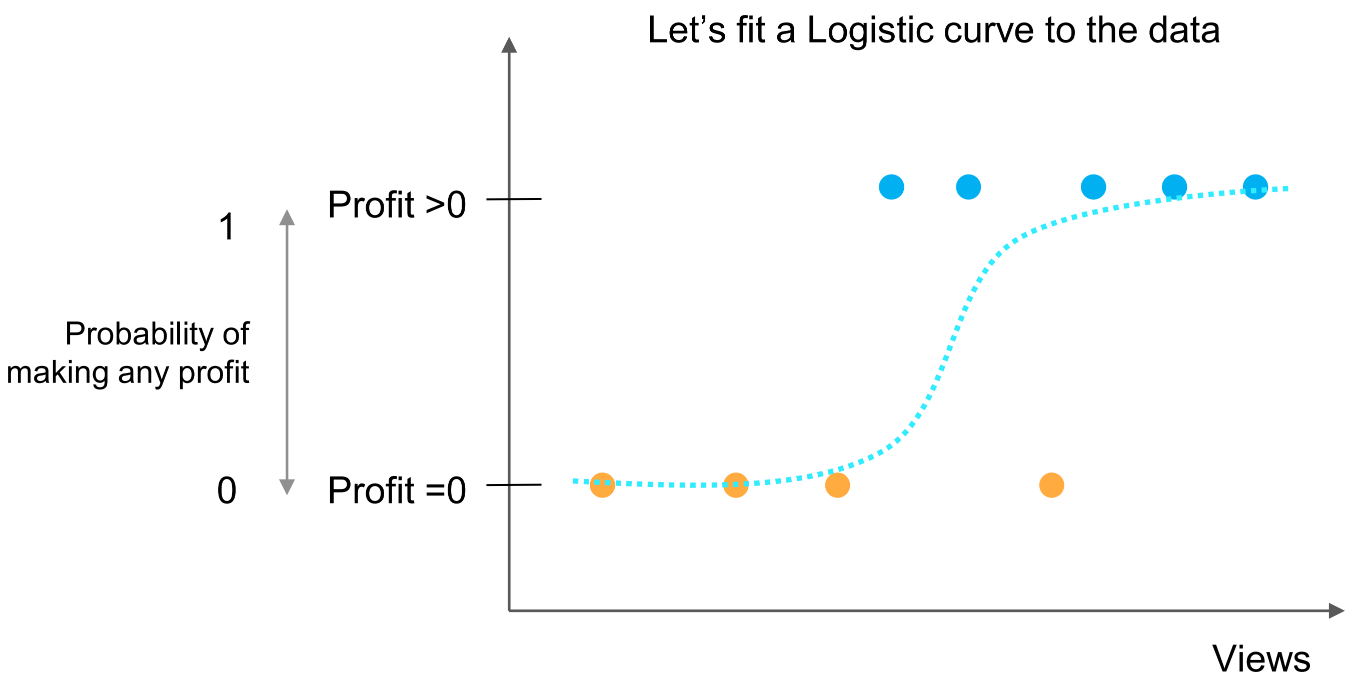
ROC (Receiver Operating Characteristic) **curve:** How to draw?



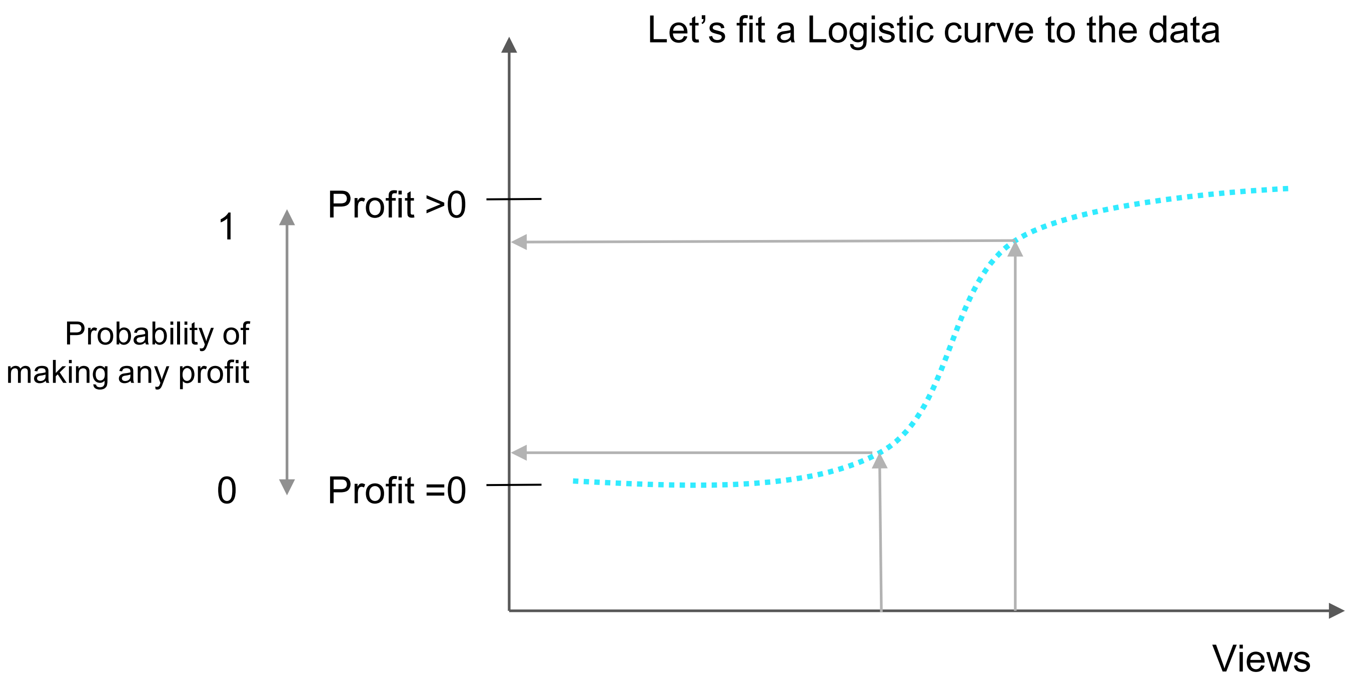
ROC (Receiver Operating Characteristic) **curve:** How to draw?



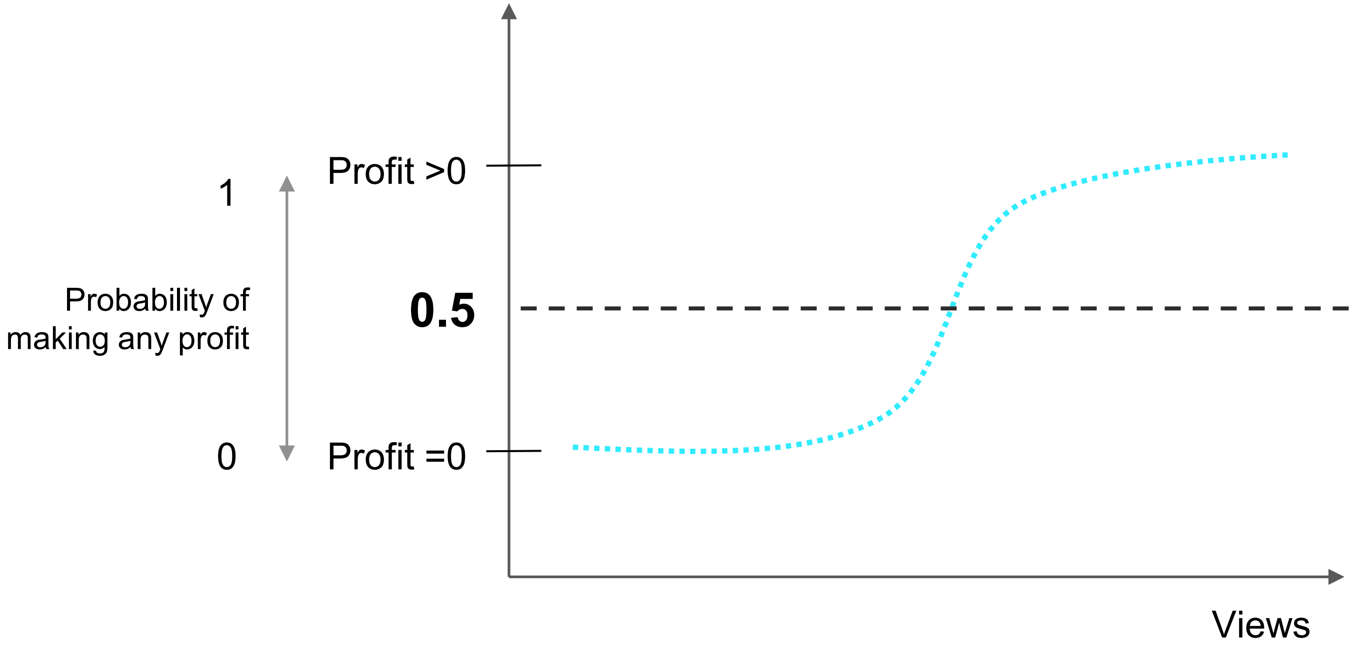
ROC (Receiver Operating Characteristic) **curve:** How to draw?



ROC (Receiver Operating Characteristic) **curve:** How to draw?

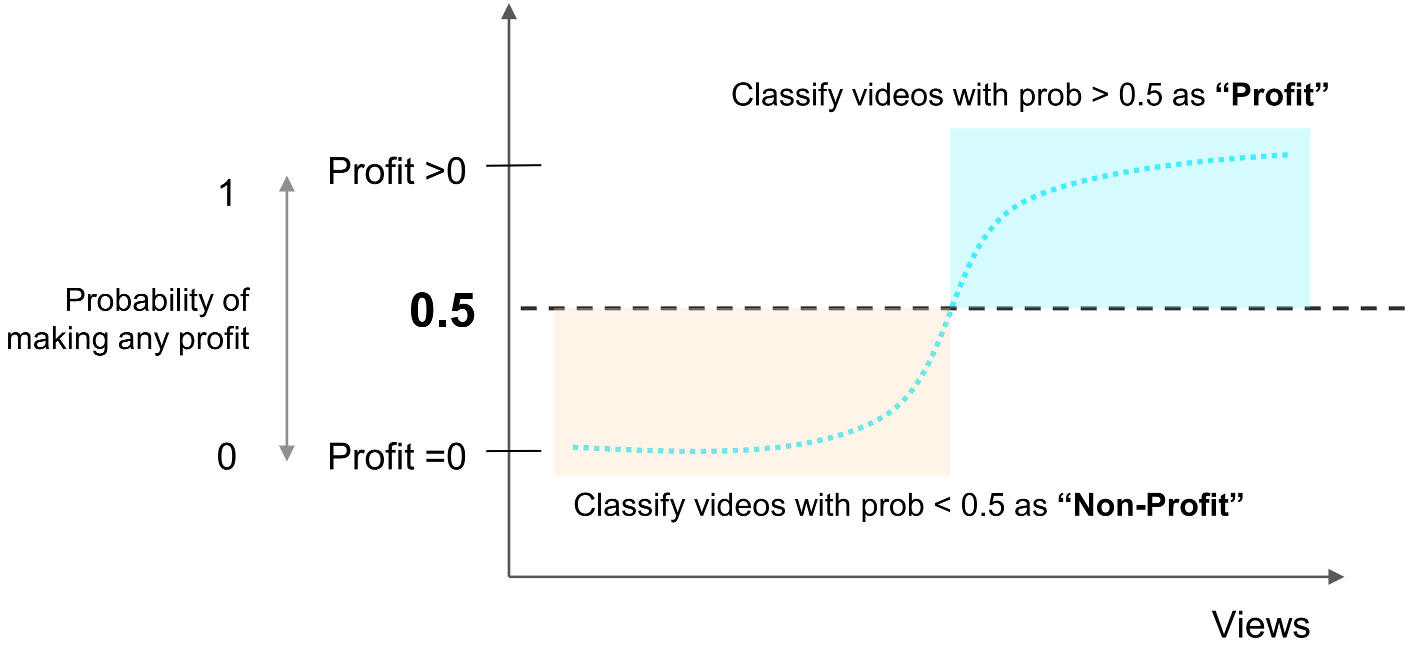


ROC (Receiver Operating Characteristic) **curve:** How to draw?



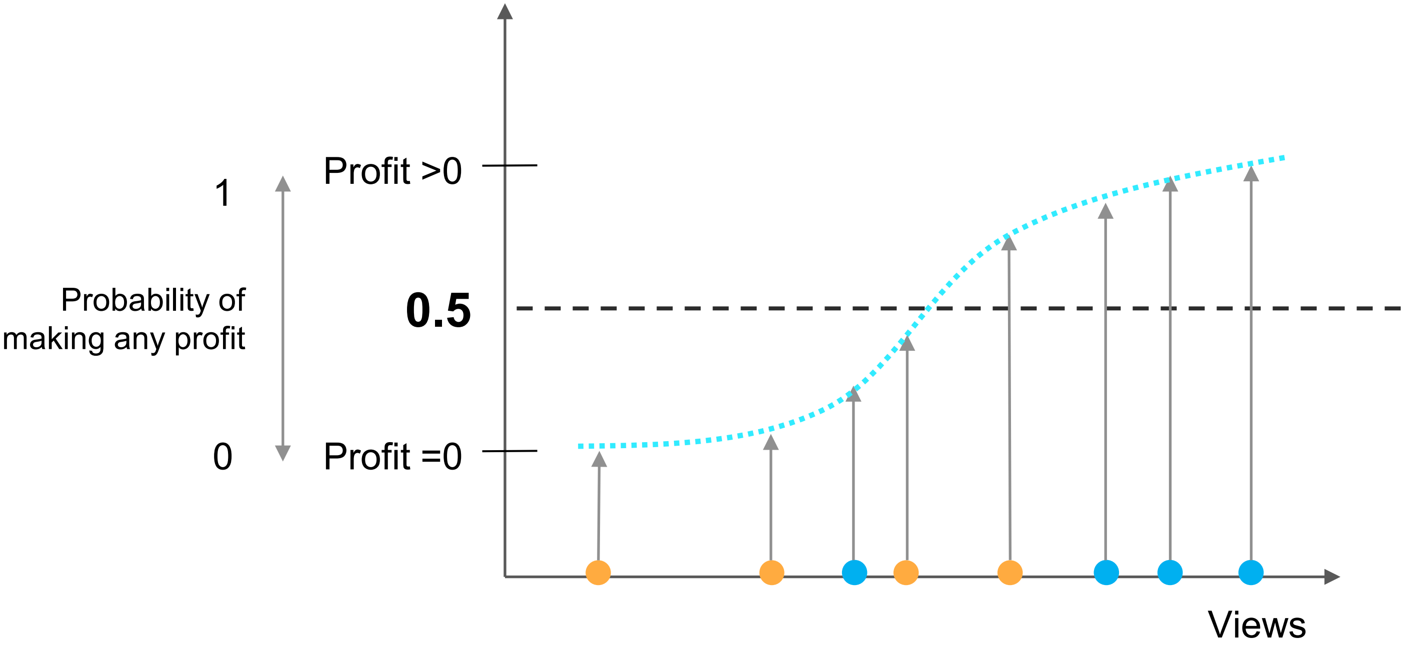
One way to classify videos is to set a **threshold** at **0.5**

ROC (Receiver Operating Characteristic) **curve:** How to draw?



One way to classify videos is to set a **threshold** at **0.5**

ROC (Receiver Operating Characteristic) **curve:** How to draw?

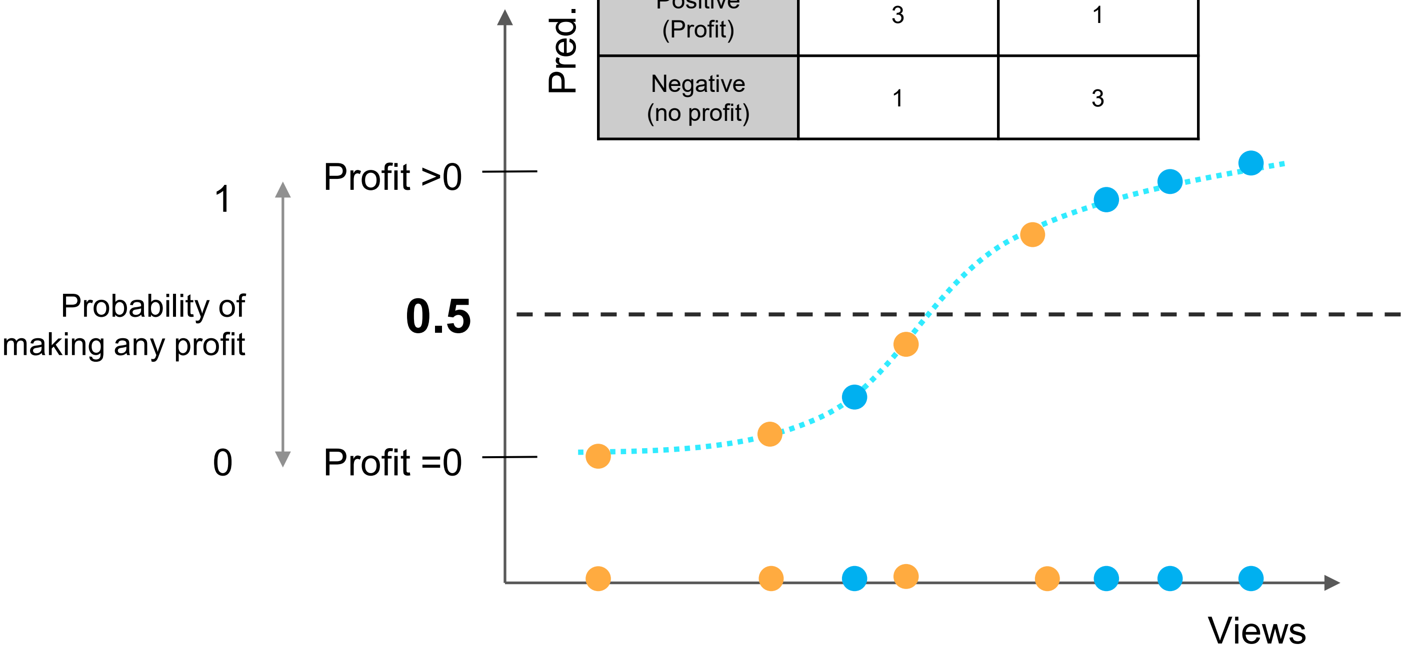


One way to classify videos is to set a **threshold** at **0.5**

ROC (Receiver Operating Characteristic) **curve:** How to draw?

Actual

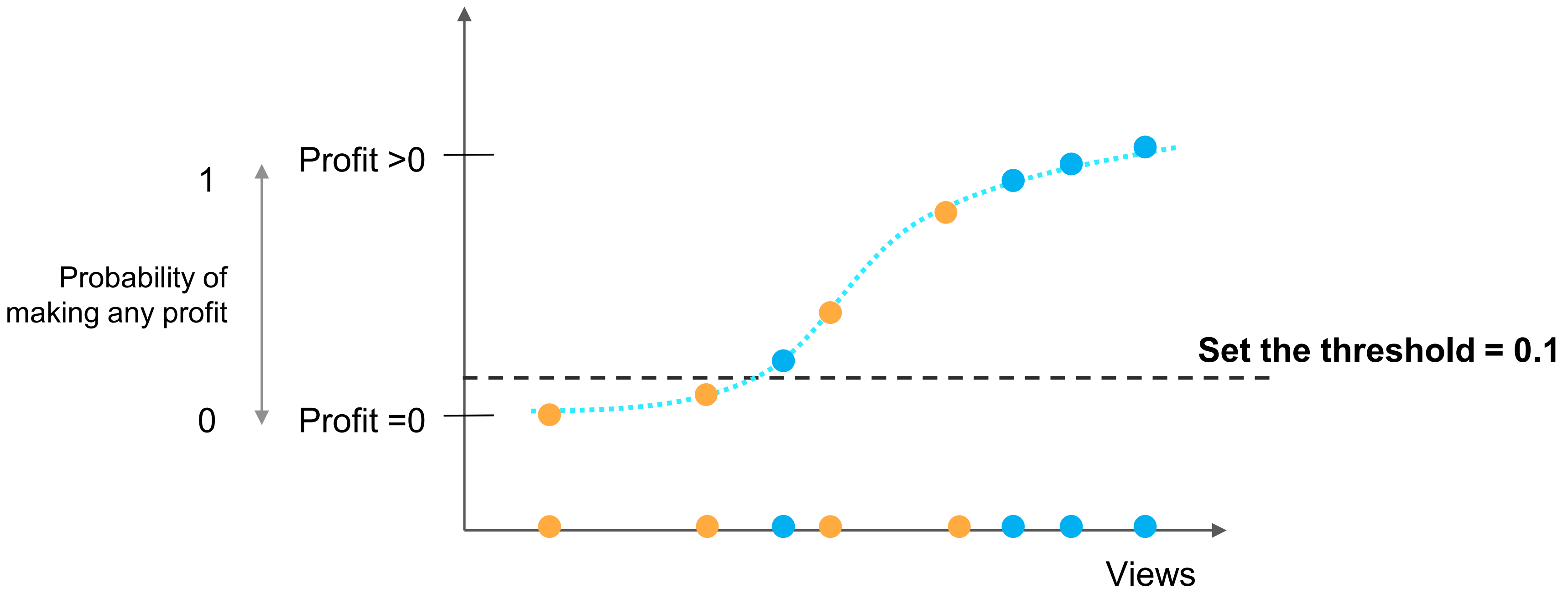
Pred.	Positive (Profit)	Negative (no profit)
Positive (Profit)	3	1
Negative (no profit)	1	3



One way to classify videos is to set a **threshold** at **0.5**

ROC (Receiver Operating Characteristic) **curve:** How to draw?

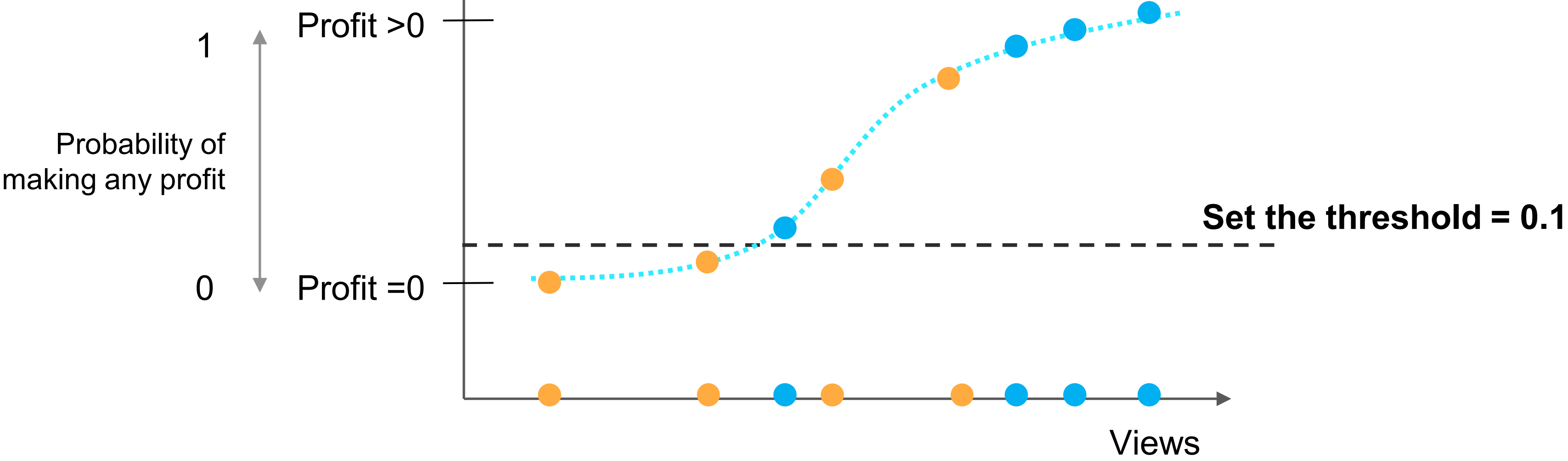
What happens when changing the threshold?



ROC (Receiver Operating Characteristic) **curve:** How to draw?

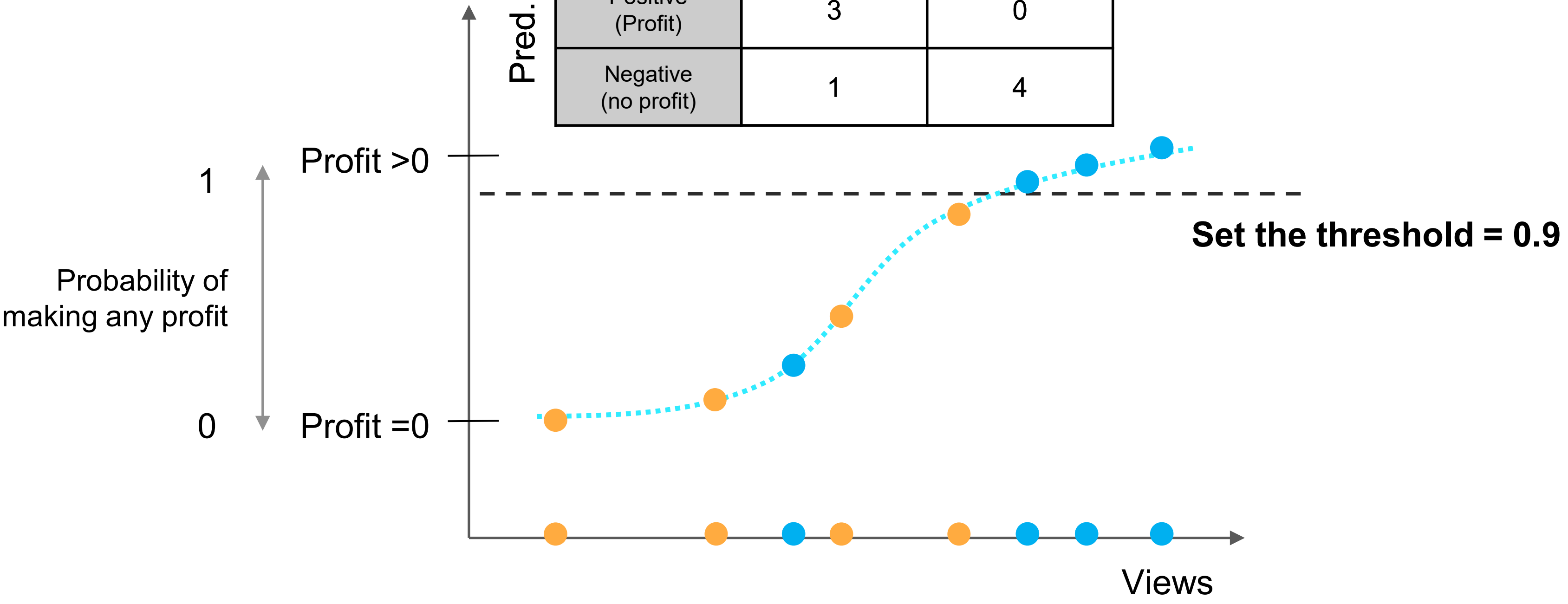
Actual

Pred.	Positive (Profit)	Negative (no profit)
Positive (Profit)	4	2
Negative (no profit)	0	2



ROC (Receiver Operating Characteristic) **curve:** How to draw?

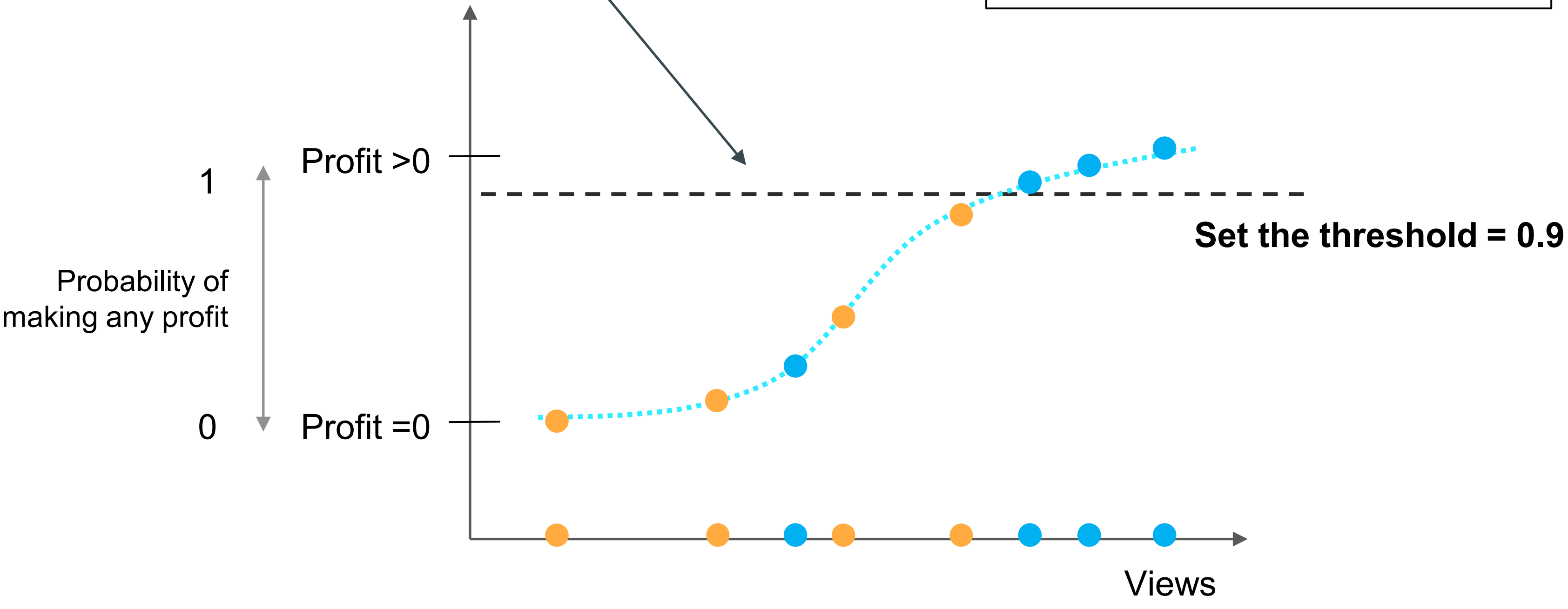
		Actual	
		Positive (Profit)	Negative (no profit)
Pred.	Positive (Profit)	3	0
	Negative (no profit)	1	4



ROC (Receiver Operating Characteristic) **curve:** How to draw?

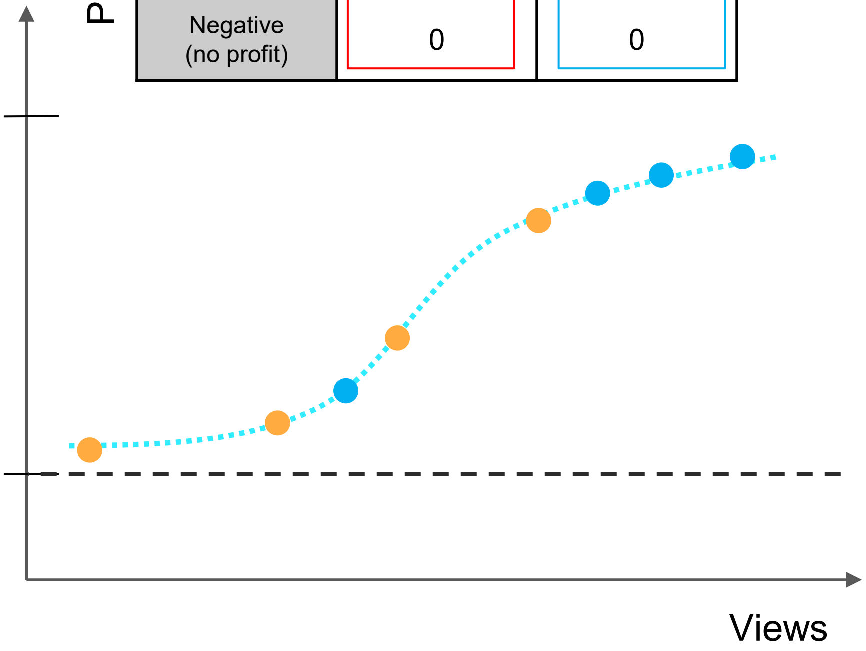
Whenever the threshold changes, we get a different confusion matrix

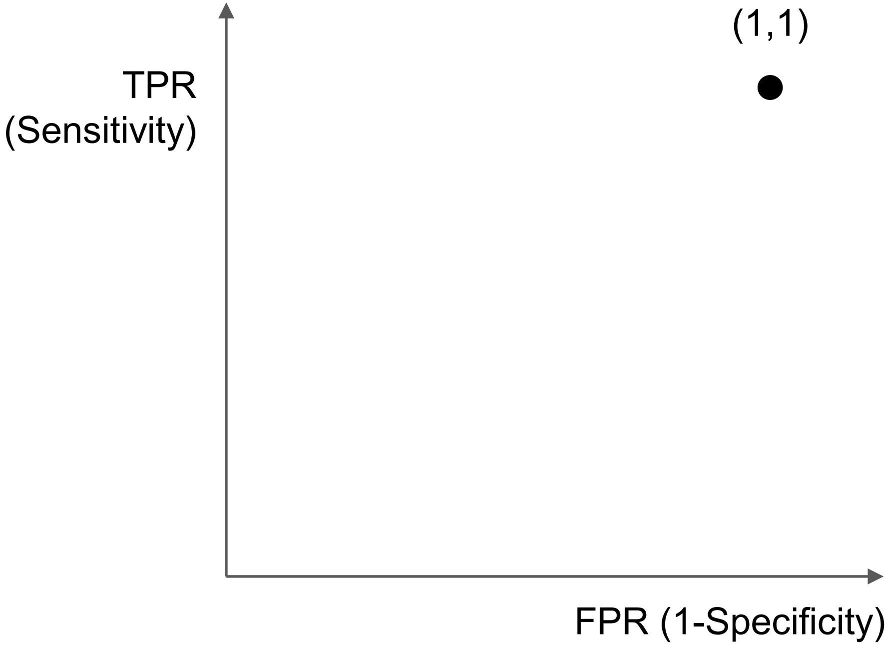
Receiver Operating Characteristic graphs provide us a simple way to summarize all of the information



ROC (Receiver Operating Characteristic) **curve:** How to draw?

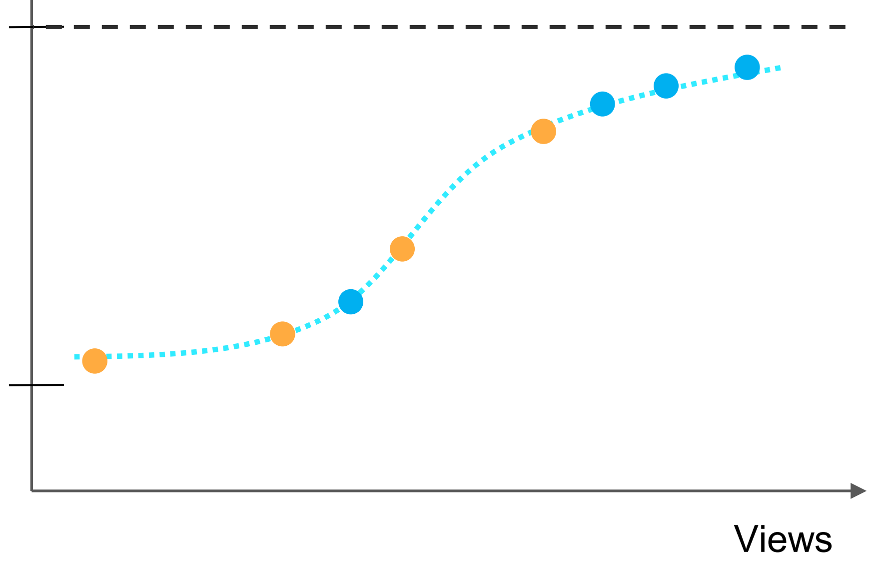
		Actual	
		Positive (Profit)	Negative (no profit)
Pred.	Positive (Profit)	4	4
	Negative (no profit)	0	0

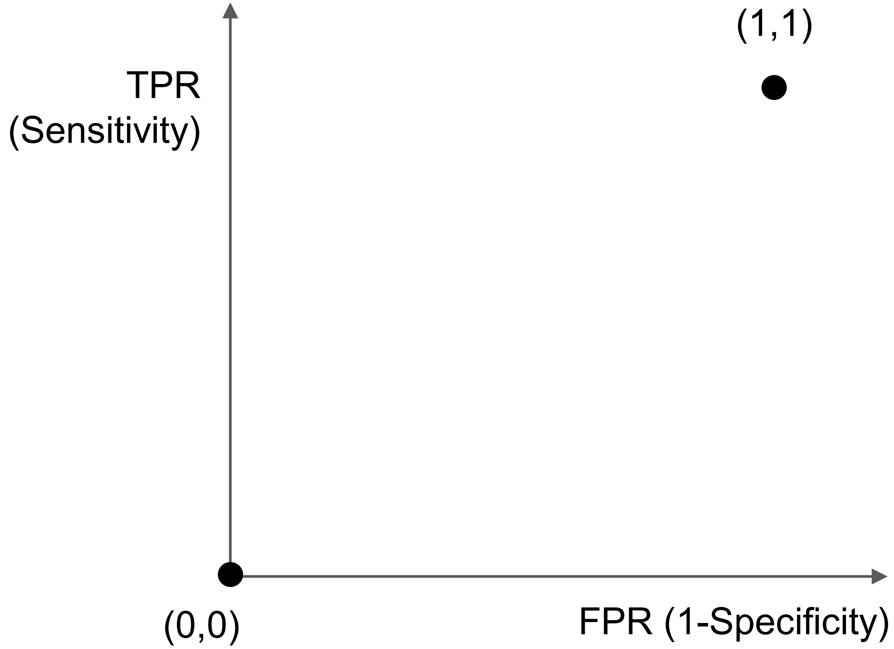




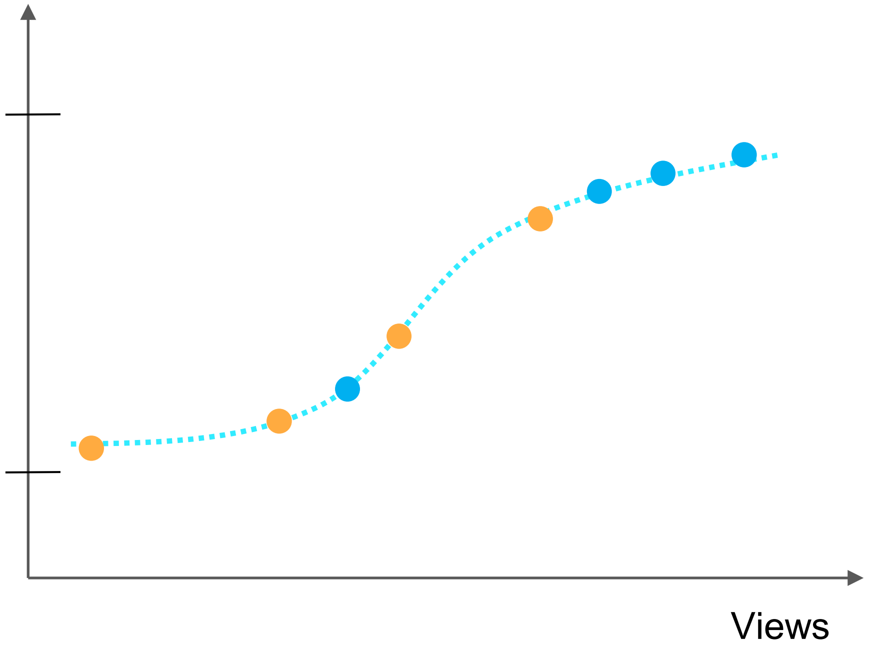
ROC (Receiver Operating Characteristic) **curve:** How to draw?

	Actual: Positive (Profit)	Actual: Negative (no profit)
Pred. Positive (Profit)	0	0
Pred. Negative (no profit)	4	4

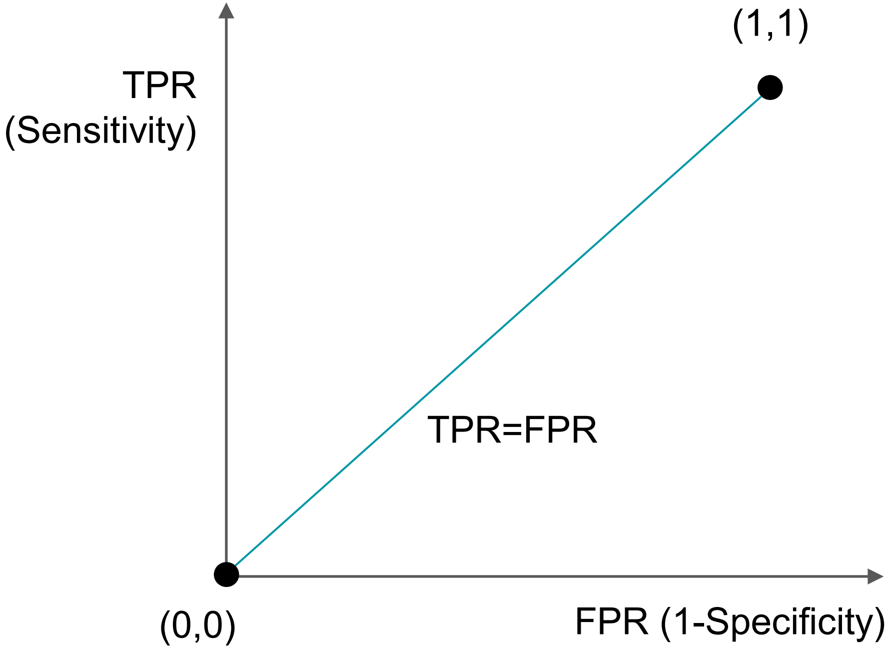




ROC (Receiver Operating Characteristic) **curve:** How to draw?

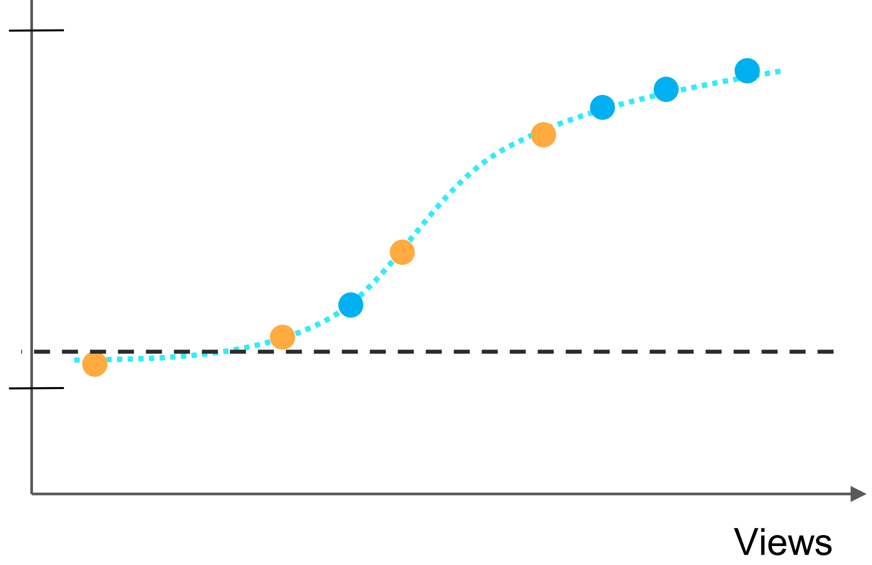


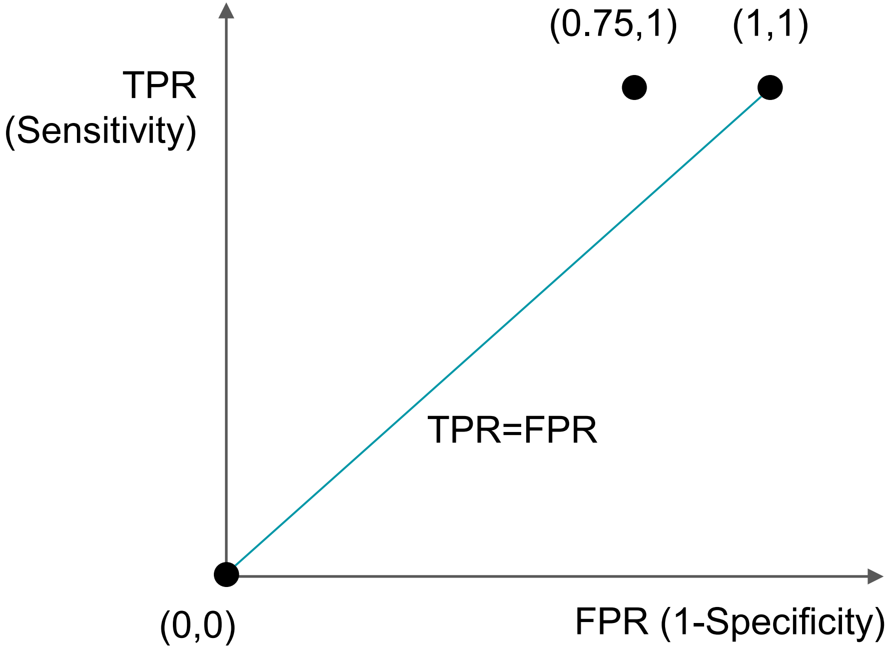
Points on the line → Proportion of correctly classified (TPR) is the same as the proportion of incorrectly classified (FPR) → Random classifier



ROC (Receiver Operating Characteristic) **curve:** How to draw?

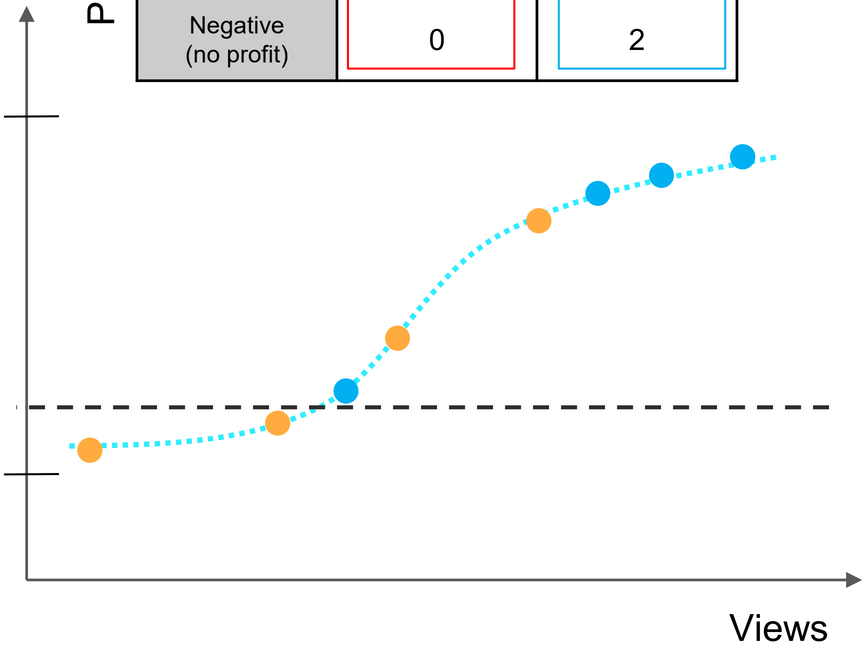
Pred. \ Actual	Positive (Profit)	Negative (no profit)
Positive (Profit)	4	3
Negative (no profit)	0	1

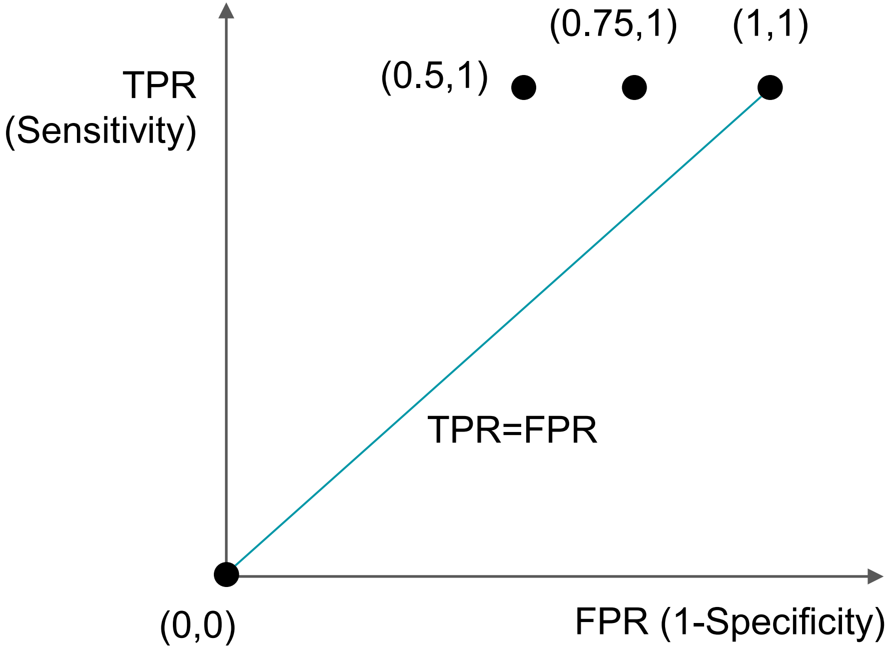




ROC (Receiver Operating Characteristic) **curve:** How to draw?

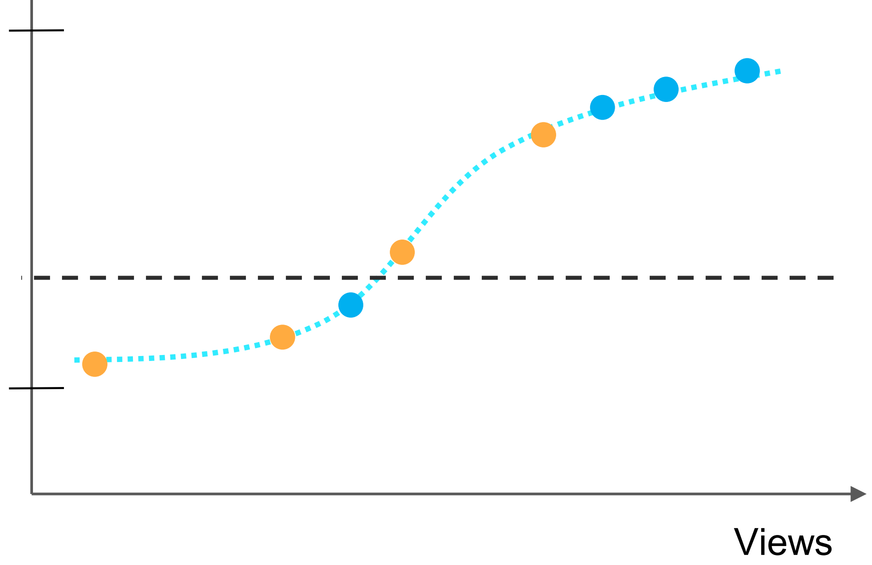
	Actual: Positive (Profit)	Actual: Negative (no profit)
Pred. Positive (Profit)	4	2
Pred. Negative (no profit)	0	2

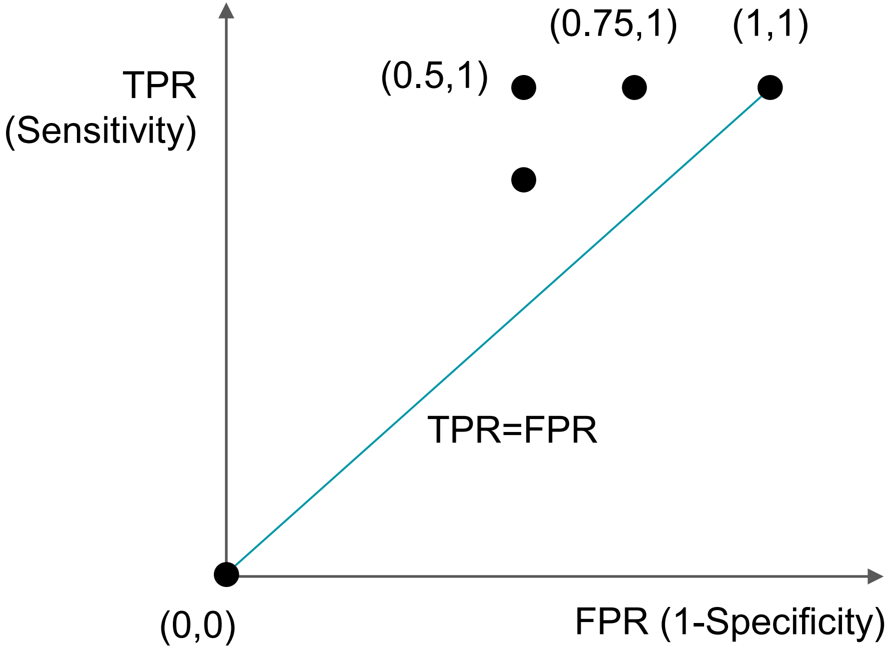




ROC (Receiver Operating Characteristic) **curve:** How to draw?

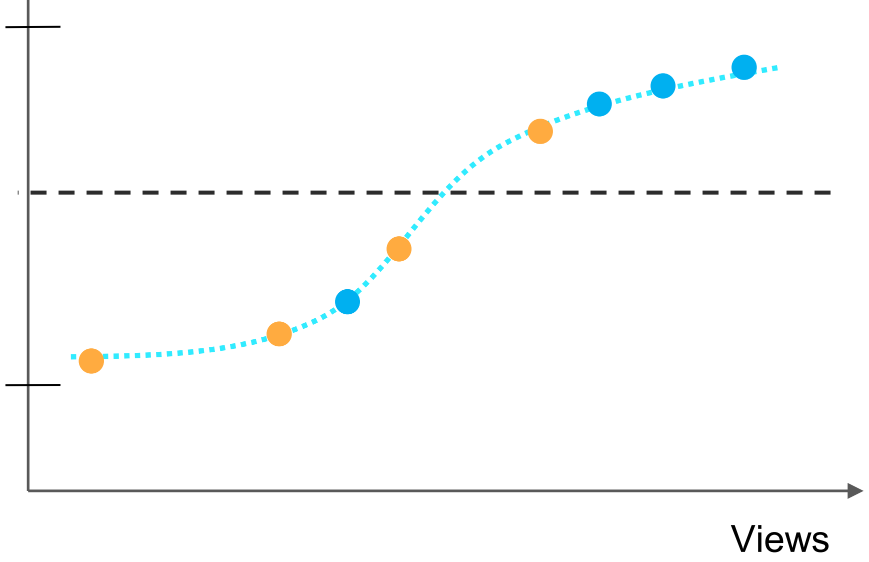
Pred. \ Actual	Positive (Profit)	Negative (no profit)
Positive (Profit)	3	2
Negative (no profit)	1	2

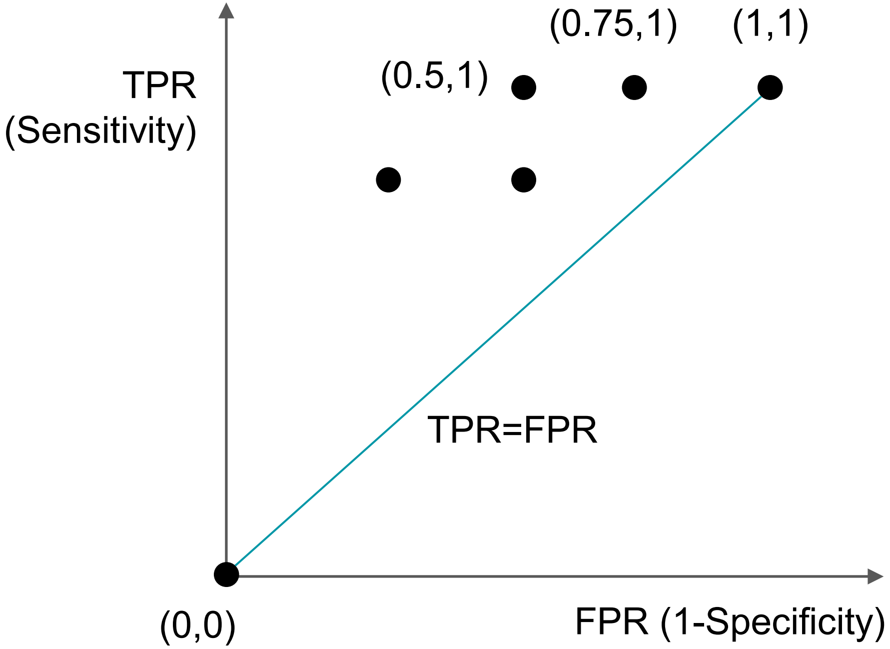




ROC (Receiver Operating Characteristic) **curve:** How to draw?

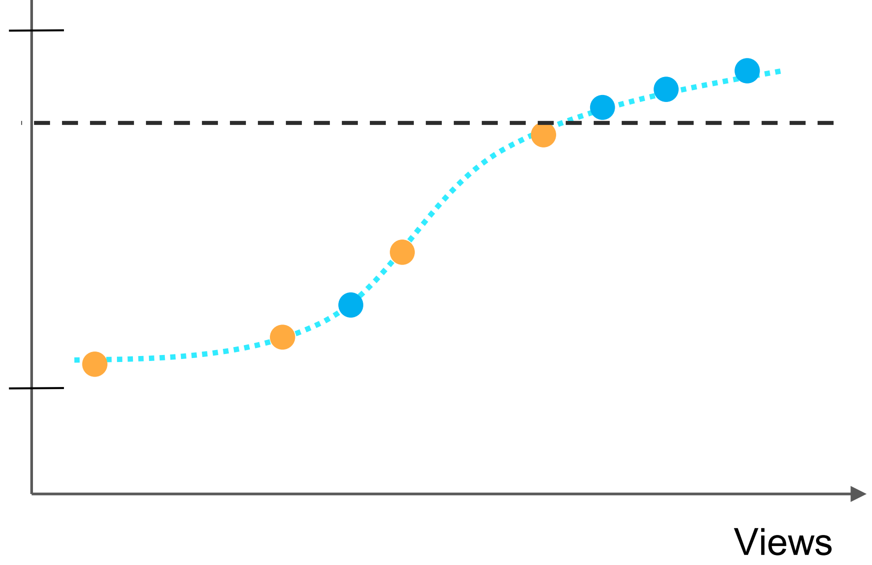
		Actual	
		Positive (Profit)	Negative (no profit)
Pred.	Positive (Profit)	3	1
	Negative (no profit)	1	3

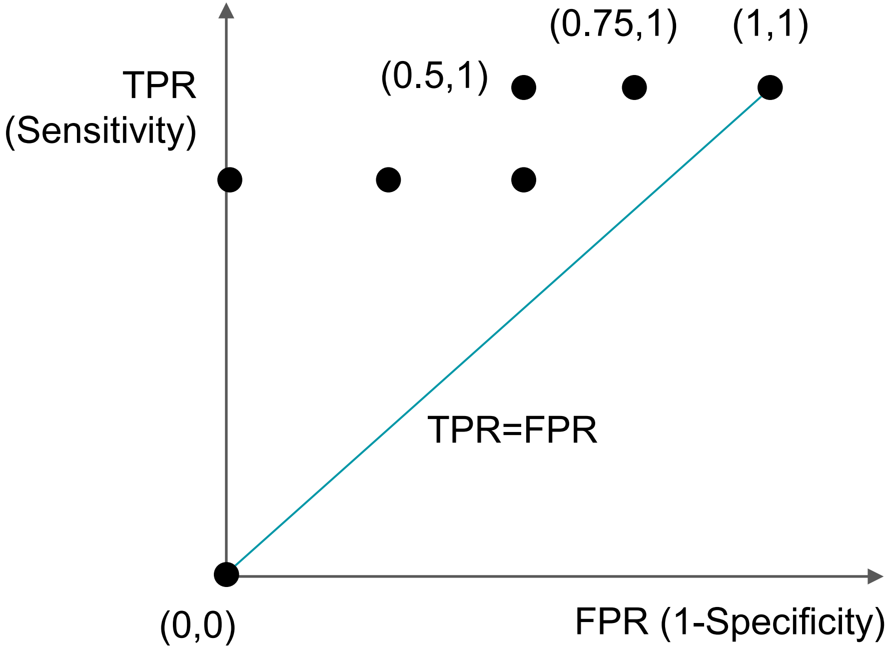




ROC (Receiver Operating Characteristic) **curve:** How to draw?

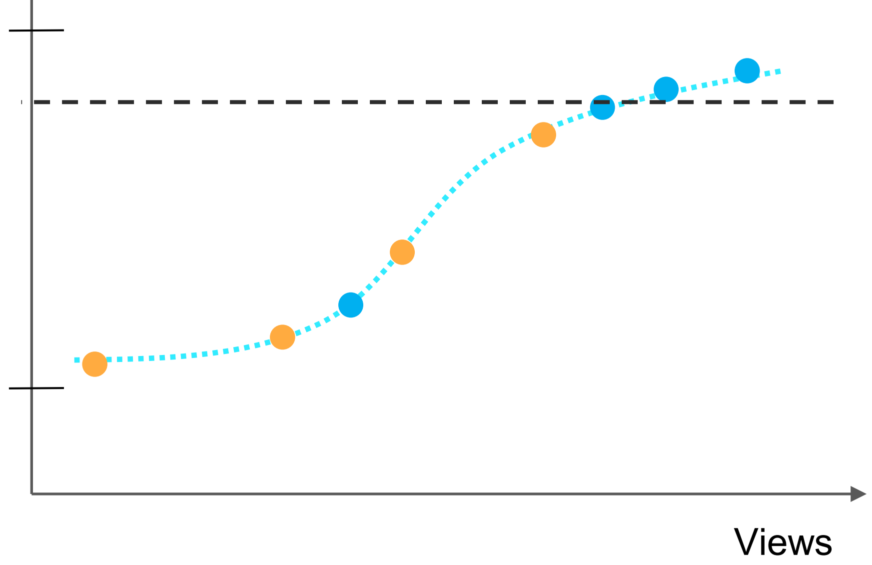
Pred. \ Actual	Positive (Profit)	Negative (no profit)
Positive (Profit)	3	0
Negative (no profit)	1	4

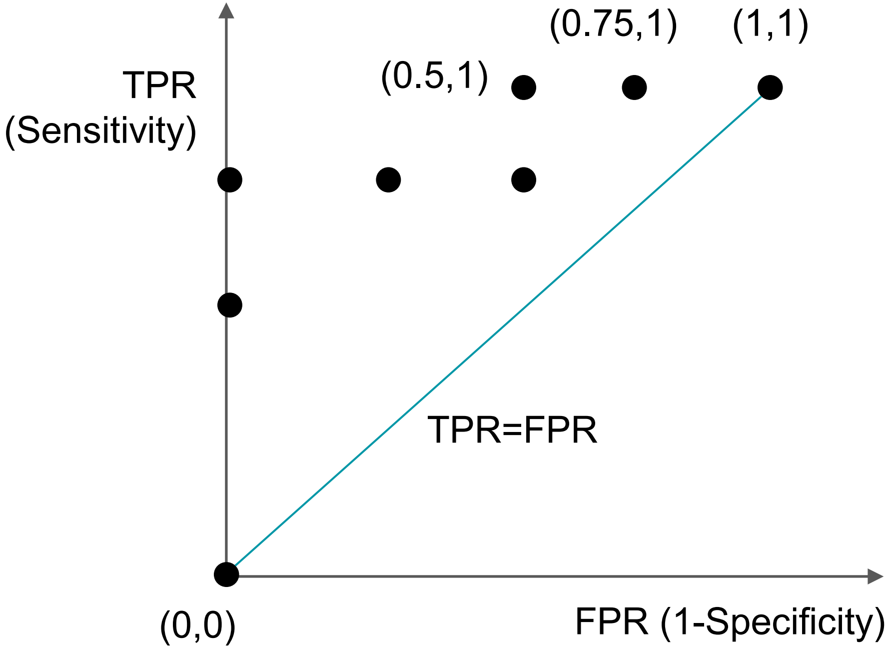




ROC (Receiver Operating Characteristic) **curve:** How to draw?

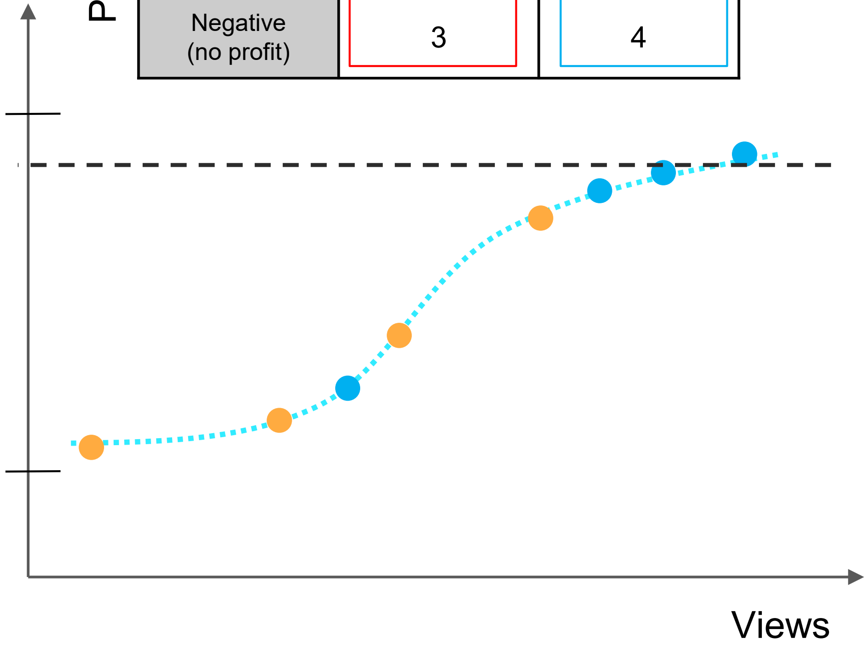
Pred. \ Actual	Positive (Profit)	Negative (no profit)
Positive (Profit)	2	0
Negative (no profit)	2	4

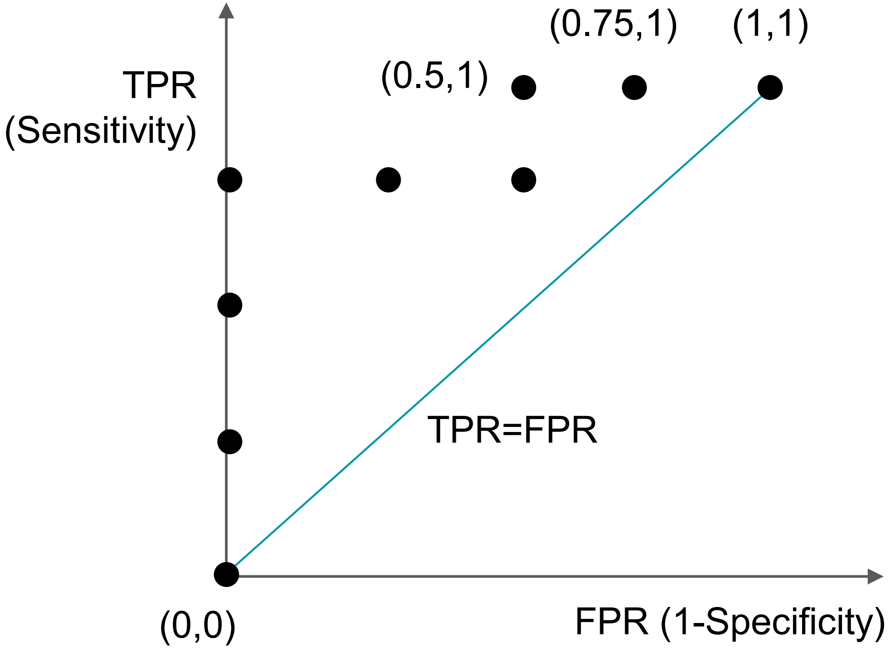




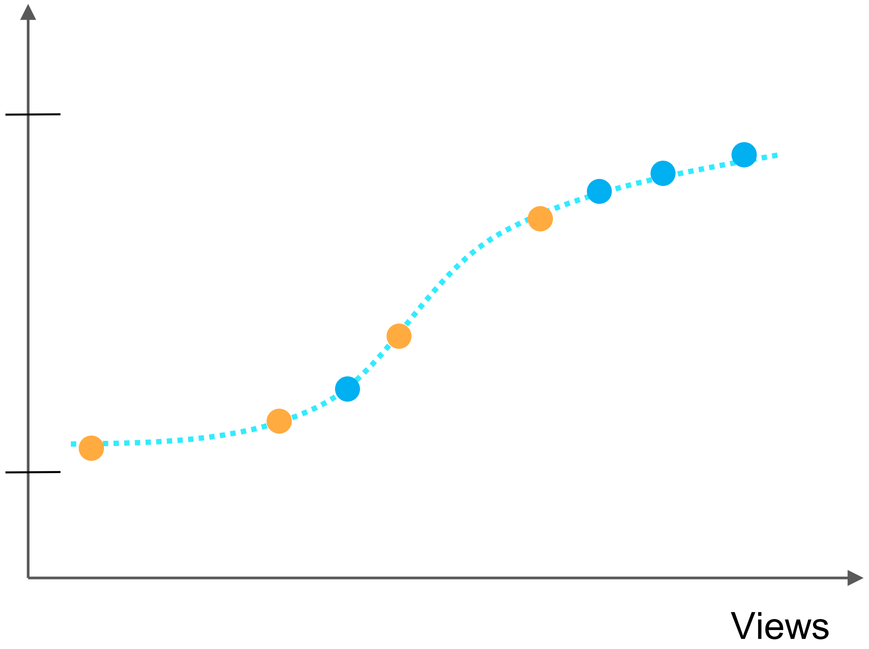
ROC (Receiver Operating Characteristic) **curve:** How to draw?

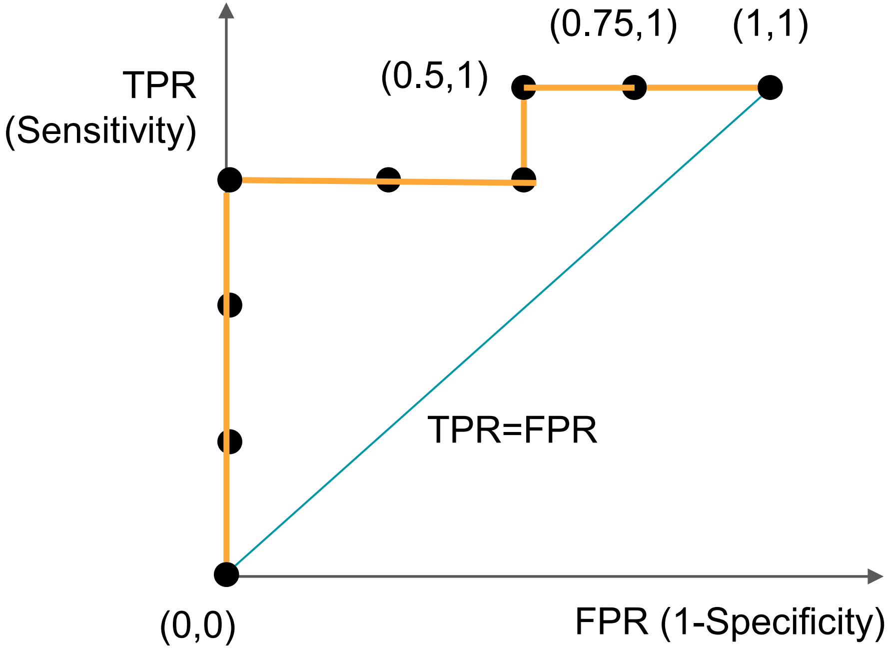
		Actual	
		Positive (Profit)	Negative (no profit)
Pred.	Positive (Profit)	1	0
	Negative (no profit)	3	4



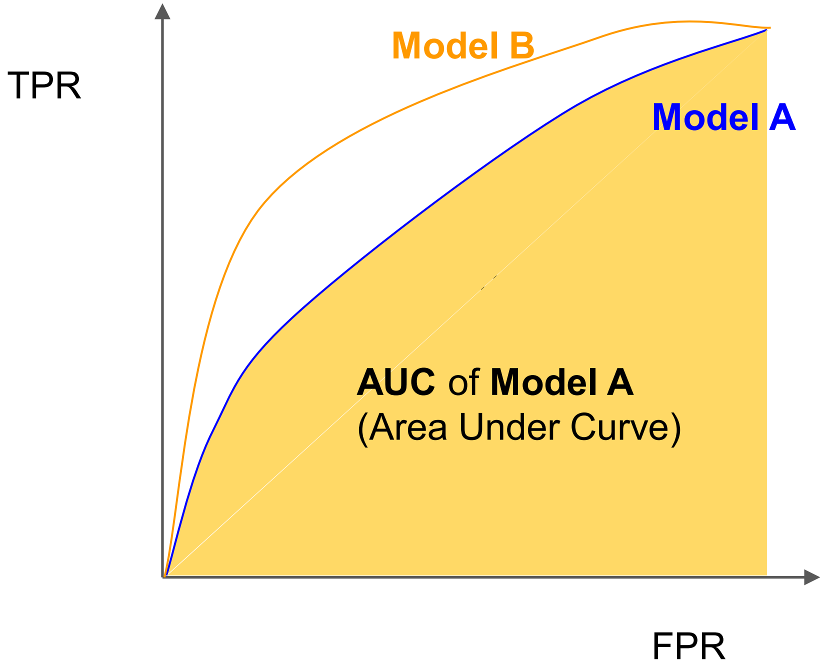


ROC (Receiver Operating Characteristic) **curve:** How to draw?





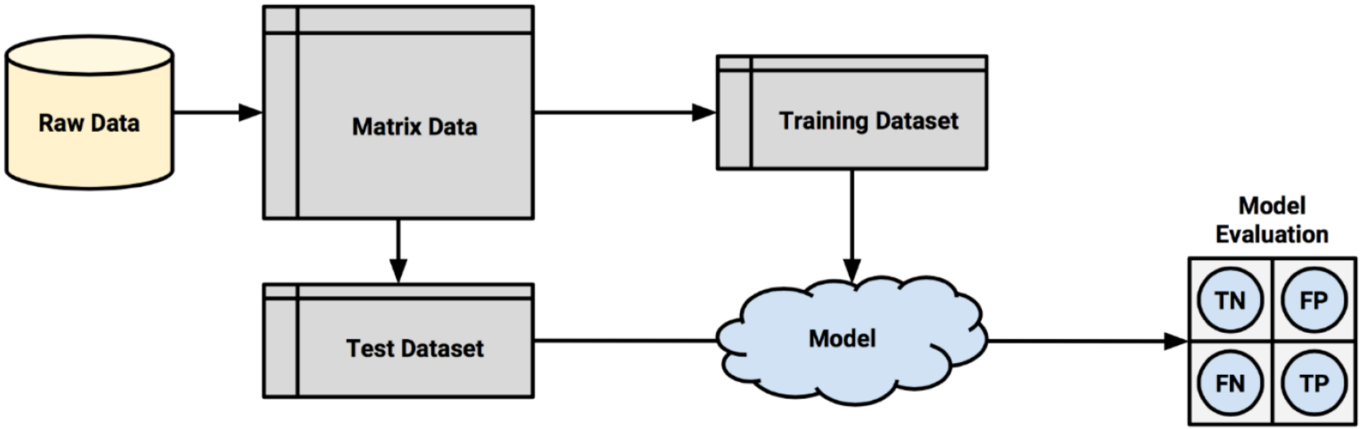
ROC (Receiver Operating Characteristic) **curve:** Model comparison



For fair evaluation

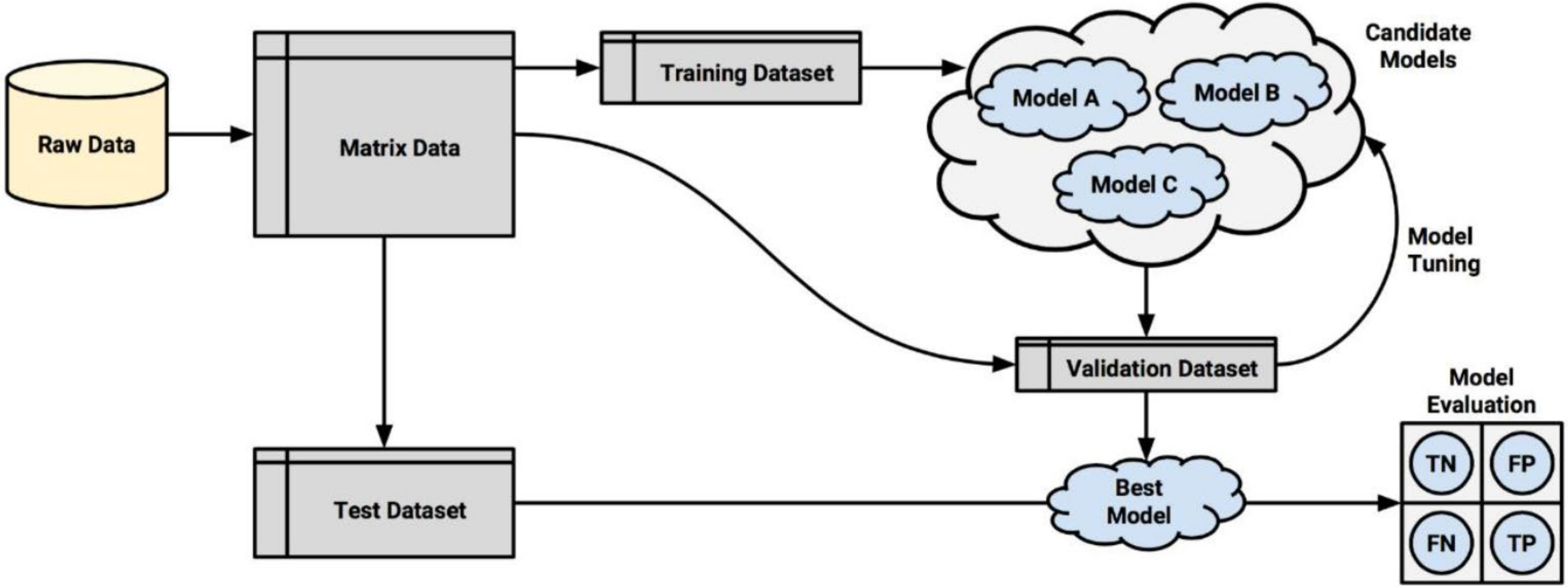
Let’s use different dataset for the test from the dataset used for training

Holdout method



For fair evaluation

Better method for model improvement



For fair evaluation

K-fold Cross-Validation