Fundamentals of Computer Science

Exercise Session 11

15.5.2020

What do we do today

- Clarify two questions

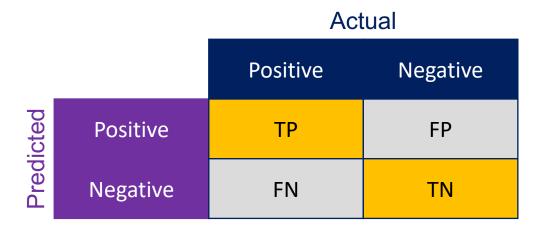- Implement Naïve Bayes Classificator

- Model performance metrics

Cofusion Matrix

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

Accuracy

Actual

	Positive	Negative
Positive	TP	FP
Negative	FN	TN

Predicted

$$Accuracy = \frac{Number\ of\ Correct\ predictions}{Total\ number\ of\ predictions\ made}$$

→ How many cases did our model predict correctly?
What if costs of miclassification are very high?
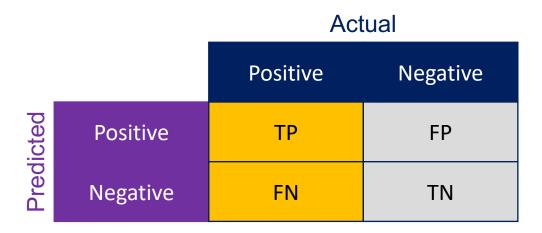Accuracy metric is not sufficient.

Precision

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

→ What ratio of the items predicted as positive are actually positive?
e.g. Detecting spam emails we would be happy with high precision

Recall

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

→ What ratio of the actual positives did our model predict as positive?
e.g. Identifying sick patients this ratio is crucial

Precision or Recall?

Your boss has no clue about the differences!

F1 measure

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

→ Find balance between Precision & Recall if cost of misclasification is high
→ Your boss only has to look at one key figure