DLI Accelerated Data Science Teaching Kit

Lecture 14.7 - Visualizing Classification: ROC, AUC, Confusion Matrix

Visualizing Classification Performance

Confusion matrix

		Predicted class		
		Cat	Dog	Rabbit
Actual class	Cat	5	3	0
	Dog	2	3	1
	Rabbit	0	2	11

Hard to spot trends and patterns

Easier

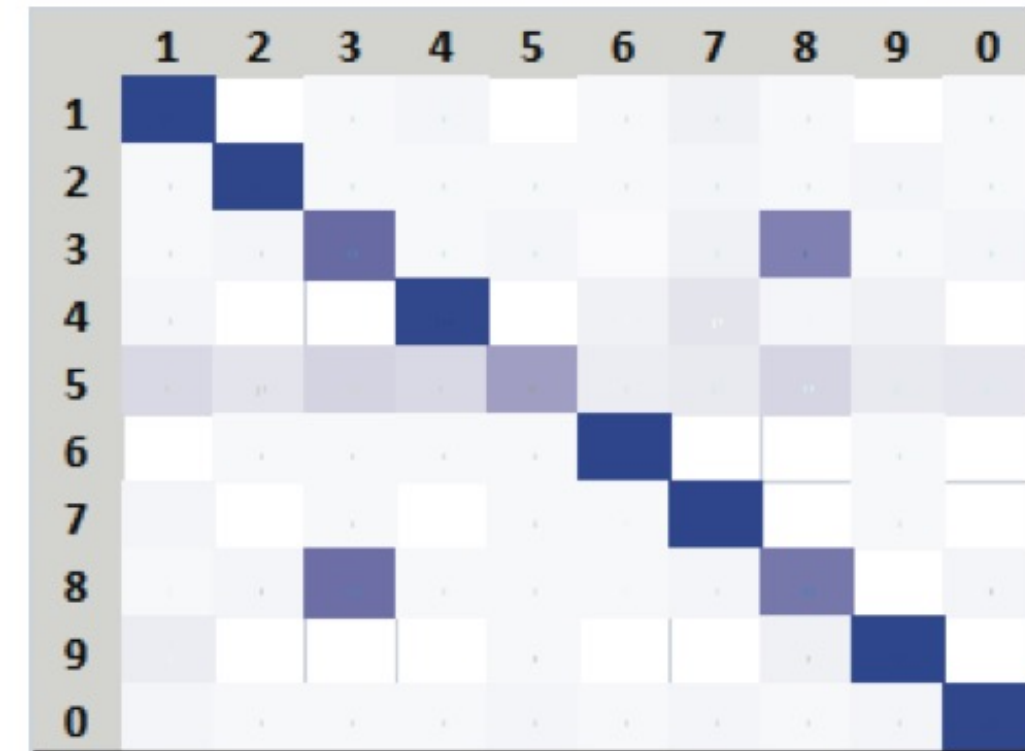|	1	2	3	4	5	6	7	8	9	0
1	91	0	1	2	0	1	3	1	0	1
2	1	89	1	1	1	1	2	1	2	1
3	1	2	48	1	2	0	3	40	1	2
4	2	0	0	83	0	3	7	2	3	0
5	10	7	12	10	30	4	5	11	5	6
6	0	1	1	1	1	95	0	0	1	0
7	2	0	1	0	1	1	94	0	1	0
8	1	2	47	1	1	1	2	43	0	2
9	4	0	0	0	1	0	0	3	92	0
0	2	1	1	1	2	1	2	1	2	87

Figure 2. Representations of confusion matrix for a handwritten digit classification task. (top) standard confusion matrix; (bottom) heat-map confusion matrix. It is much easier to identify underlying patterns in the visual representation; 3 and 8 are often misclassified as each other and 5 is misclassified as many different numbers.

http://research.microsoft.com/en-us/um/redmond/groups/cue/publications/CHI2009-EnsembleMatrix.pdf

Very Important:
Find out what "Positive" Means

		Predicted class		
		Cat	Dog	Rabbit
Actual class	**Cat**	5	3	0
	Dog	2	3	1
	Rabbit	0	2	11

5 true positives (actual cats that were correctly classified as cats)	3 false negatives (cats that were incorrectly marked as dogs)
2 false positives (dogs that were incorrectly labeled as cats)	17 true negatives (all the remaining animals, correctly classified as non-cats)

Very Important:
Find out what "Positive" Means

**Terminology and derivations
from a confusion matrix**

true positive (TP)
eqv. with hit

true negative (TN)
eqv. with correct rejection

false positive (FP)
eqv. with false alarm, Type I error

false negative (FN)
eqv. with miss, Type II error

sensitivity or **true positive rate (TPR)**
eqv. with hit rate, recall

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN}$$

specificity (SPC) or **true negative rate (TNR)**

$$SPC = \frac{TN}{N} = \frac{TN}{FP + TN}$$

precision or **positive predictive value (PPV)**

$$PPV = \frac{TP}{TP + FP}$$

recall (recall)

$$recall = \frac{TP}{TP + FN}$$

negative predictive value (NPV)

Confusion matrix. (2017, November 21). Retrieved December 04, 2017, from https://en.wikipedia.org/wiki/Confusion_matrix
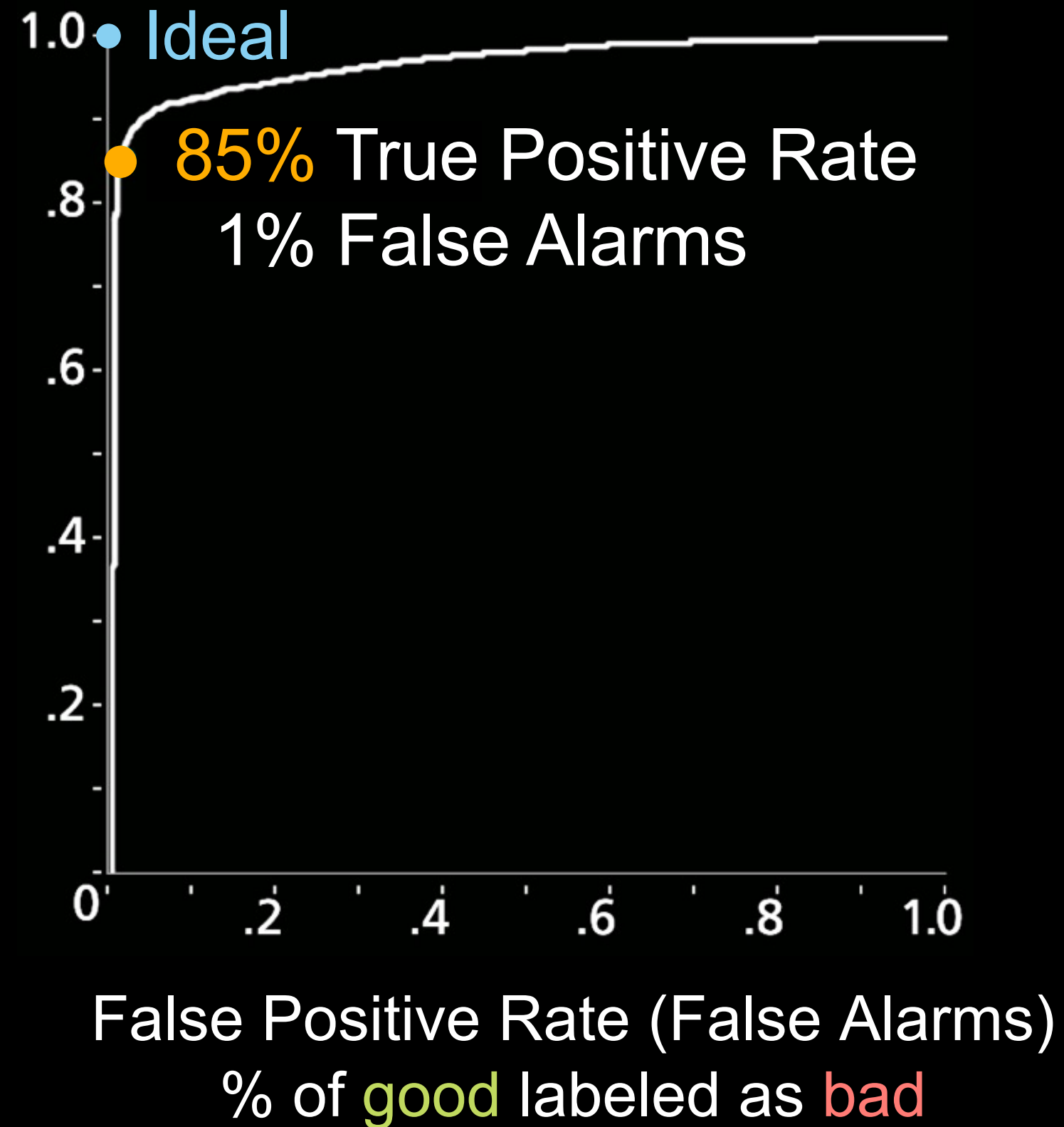
Visualizing Classification Performance using ROC curve
(Receiver Operating Characteristic)

Polonium's ROC Curve

Positive class: malware

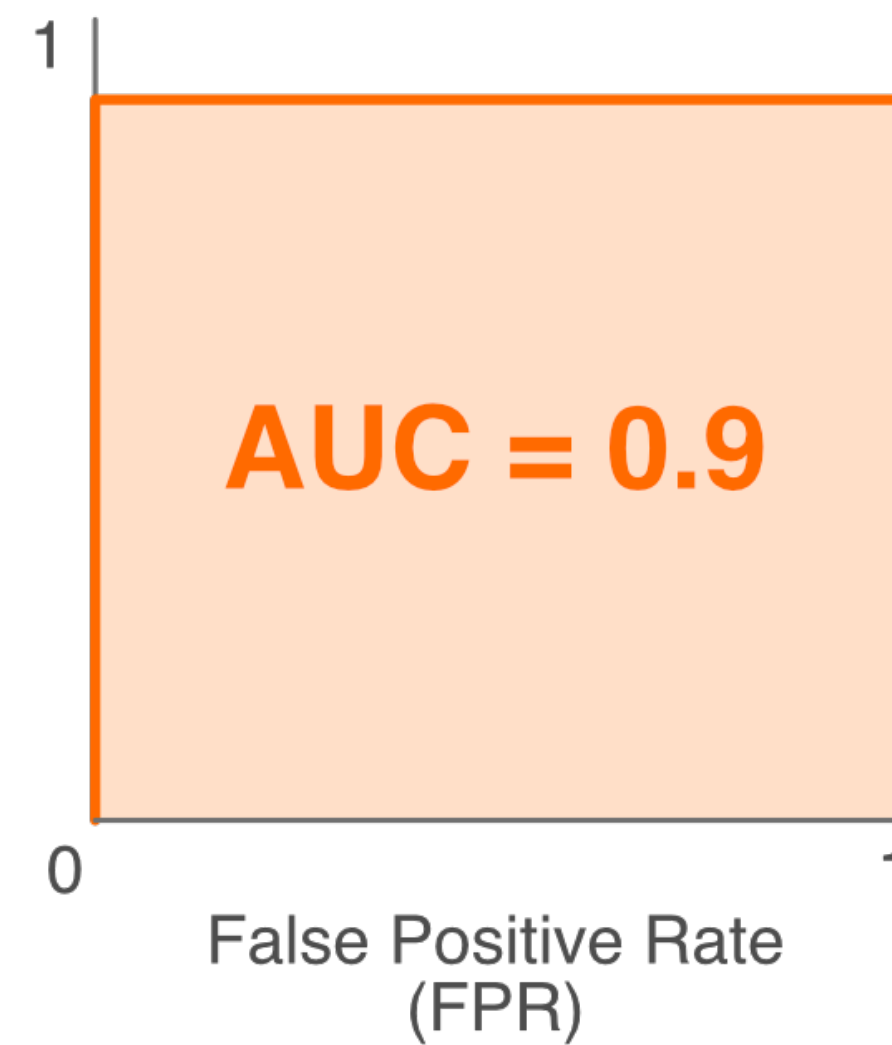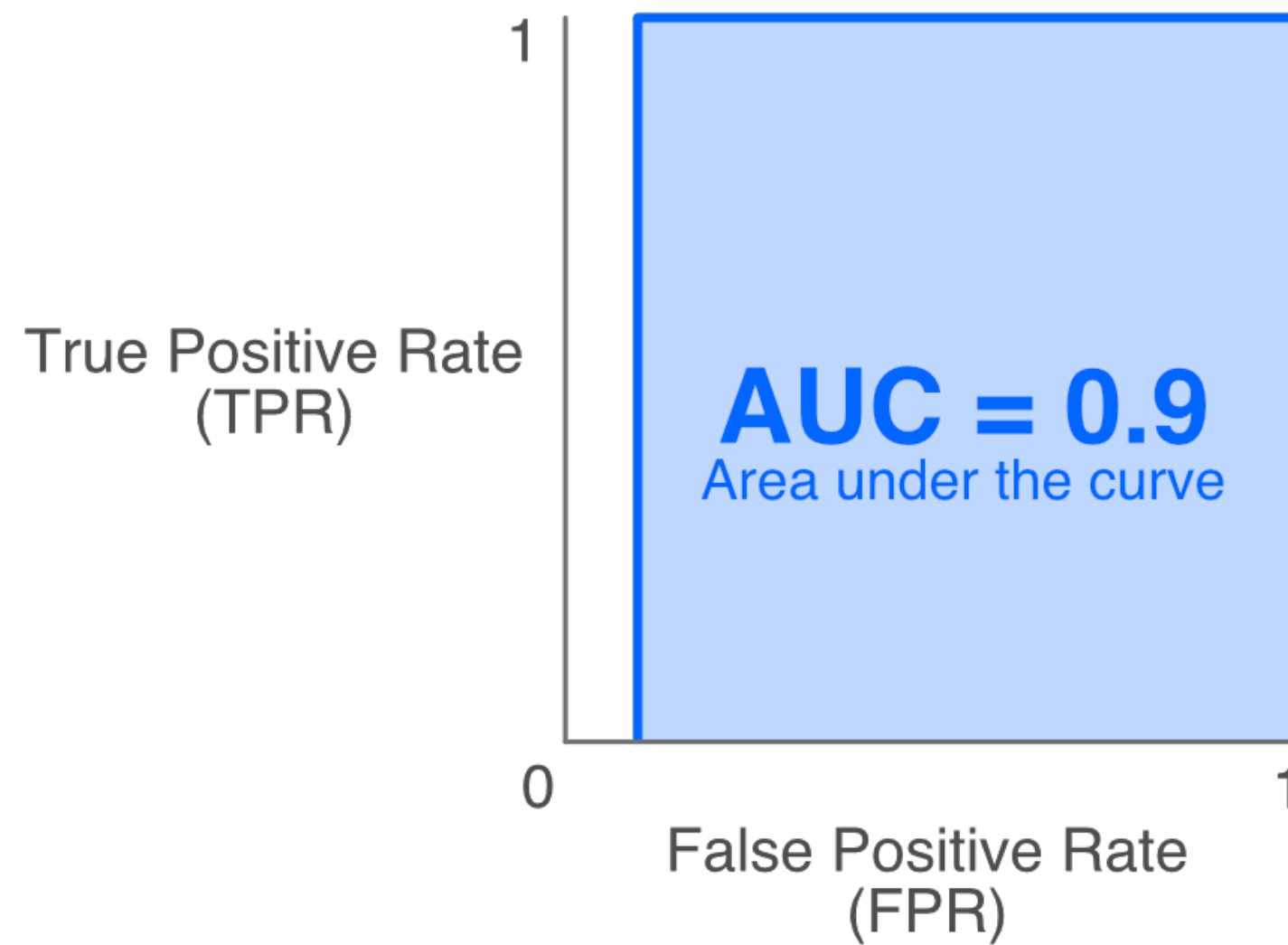Negative class: benign

True Positive Rate
% of bad correctly labeled

Ideal

85% True Positive Rate
1% False Alarms

1.0

.8

.6

.4

.2

0   .2   .4   .6   .8   1.0

False Positive Rate (False Alarms)
% of good labeled as bad

Measuring Classification Performance using AUC (Area under the curve)

If a machine learning algorithm achieves 0.9 AUC (out of 1.0).

That's a great algorithm, right?

Be Careful with AUC!

DLI Accelerated Data Science Teaching Kit

Thank You