IT496: Introduction to Data Mining

Lecture 10

Evaluation - II

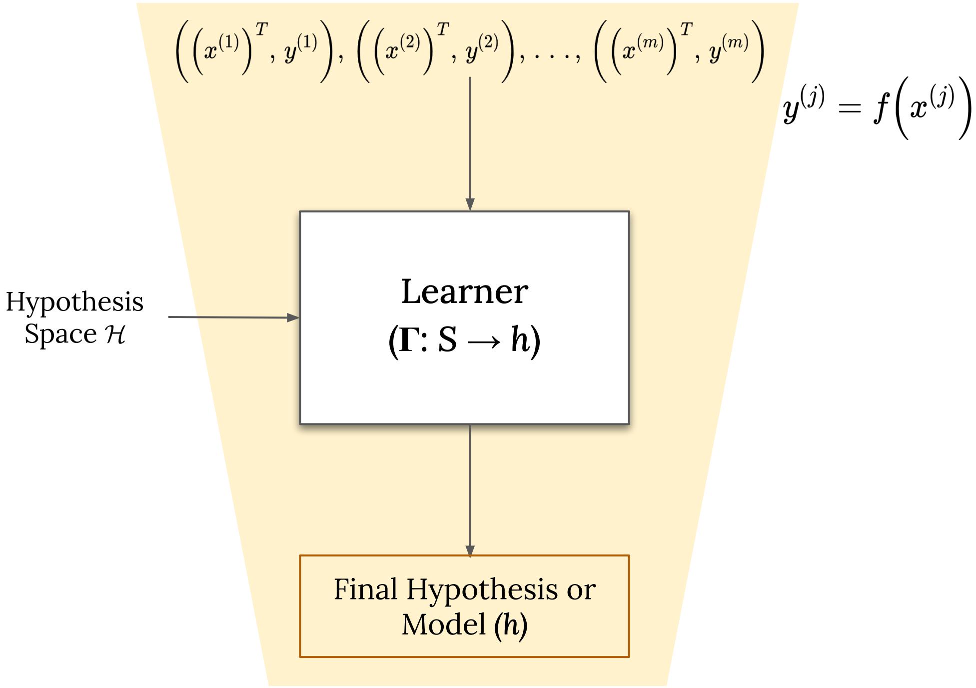
[Evaluation Metrics]

Arpit Rana

18th August 2023

Experimental Evaluation of Learning Algorithms

Given a ***representation***, ***data***, and ***a bias***, the learning algorithm returns a Final Hypothesis (h).



How to Check the Performance of Learning Algorithms?

Evaluation Metrics

Common Measures

Experimental Evaluation of Learning Algorithms

Typical Experimental Evaluation Metrics

- Error
- Accuracy
- Precision/ Recall

Measures for Regression Problems

- Mean Absolute Error

$$MAE = \frac{1}{n} \sum_{x \in S} |h(x) - y|$$

- Squared Error

$$MSE = \frac{1}{n} \sum_{x \in S} (h(x) - y)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{x \in S} (h(x) - y)^2}$$

Which one is better and why?

- Non-differentiability
- Robustness (sensitivity to outliers)
- Unit changes in MSE

Measures for Regression Problems

- Misclassification Rate (a.k.a. Error Rate)

$$MR = \frac{1}{n} \sum_{x \in S} \delta(h(x), y)$$

Where, $$\delta(h(x), y) = \begin{cases} 1 & h(x) \neq y \\ 0 & otherwise \end{cases}$$

Measures for Classification Problems

Confusion Matrix		True Class (Actual)		
		Positive	Negative	Total
Hypothesized Class (Predicted)	Positive	True Positive (TP)	False Positive (FP)	P’
	Negative	False Negative (FN)	True Negative (TN)	N’
Total		P	N	P + N

$$accuracy = \frac{TP + TN}{P + N}$$

$$\begin{aligned} error\ rate &= 1 - accuracy \\ &= \frac{FP + FN}{P + N} \end{aligned}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} = \frac{TP}{P'}$$

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

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

Measures for Classification Problems

Confusion Matrix		True Class (Actual)		
		Positive	Negative	Total
Hypothesized Class (Predicted)	Positive	True Positive (TP)	False Positive (FP)	P’
	Negative	False Negative (FN)	True Negative (TN)	N’
Total		P	N	P + N

***F* measure**: weighted harmonic mean of *precision* and *recall*.

$$F = \frac{1}{\alpha . \frac{1}{Precision} + (1-\alpha) . \frac{1}{Recall}}$$

$$= \frac{\left(\beta^2 + 1\right) Precision . Recall}{\beta^2 . Precision + Recall}$$

$$where, \quad \beta^2 = \frac{1-\alpha}{\alpha}$$

$\alpha \in [0, 1]$ and $\beta \in [0, \infty]$

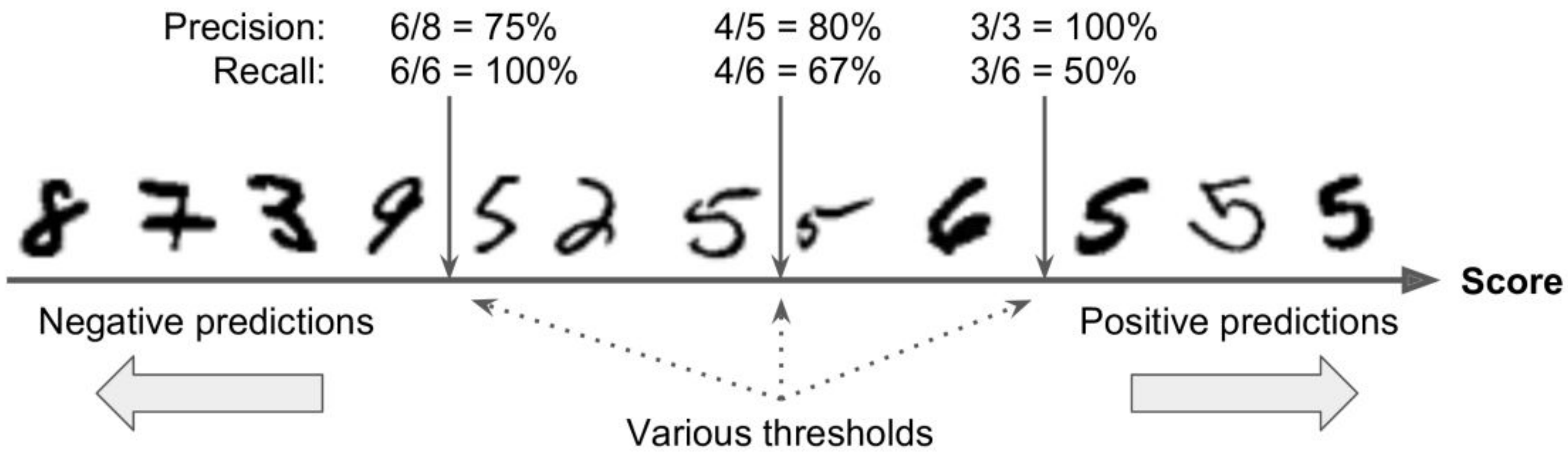
For α **= ½,** β **= 1,** F measure will be <u>*balanced*</u> and is known as F_1 *measure*.

Measures for Classification Problems

What metric would you use to measure the performance of the following classifiers.

- A classifier to detect videos that are <u>*safe for kids*</u>.
- A classifier to detect <u>*shoplifters*</u> in surveillance images.

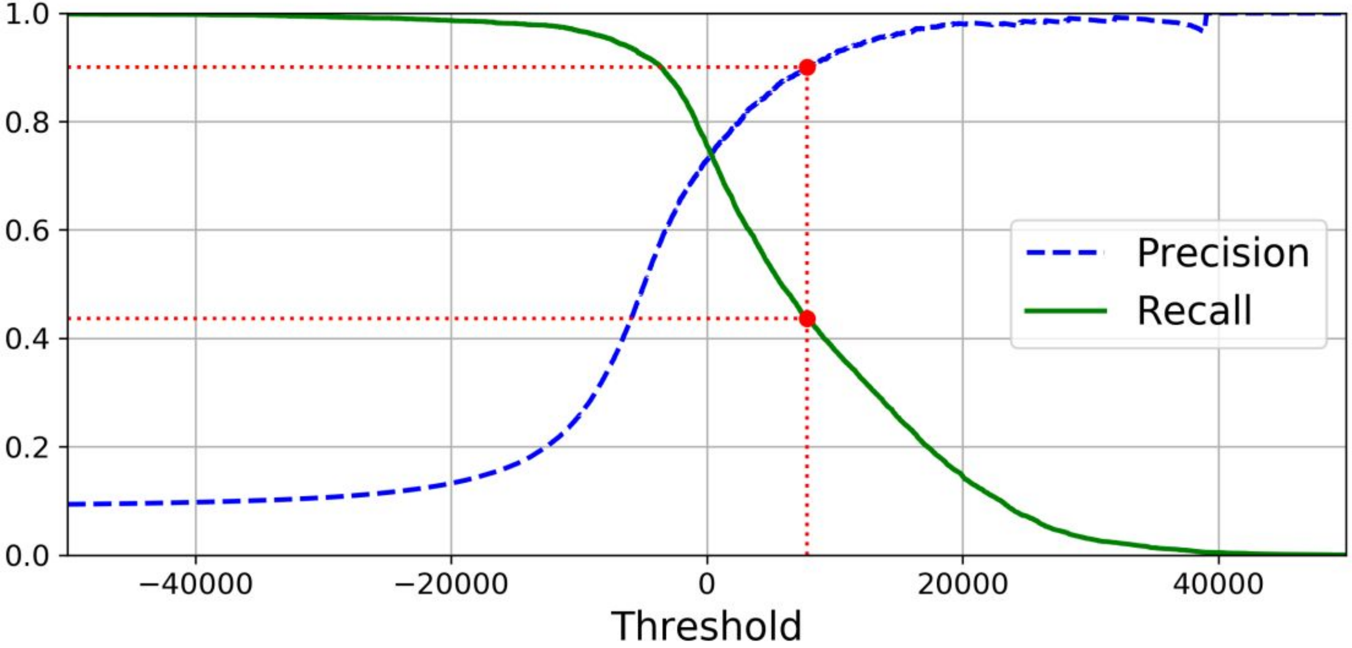
Precision/Recall Trade-off



- Images are ranked by their classifier (*whether the image is 5 or not*) score.
- Those above the chosen decision threshold are considered positive.
- The higher the threshold, the lower the recall, but (in general) the higher the precision.

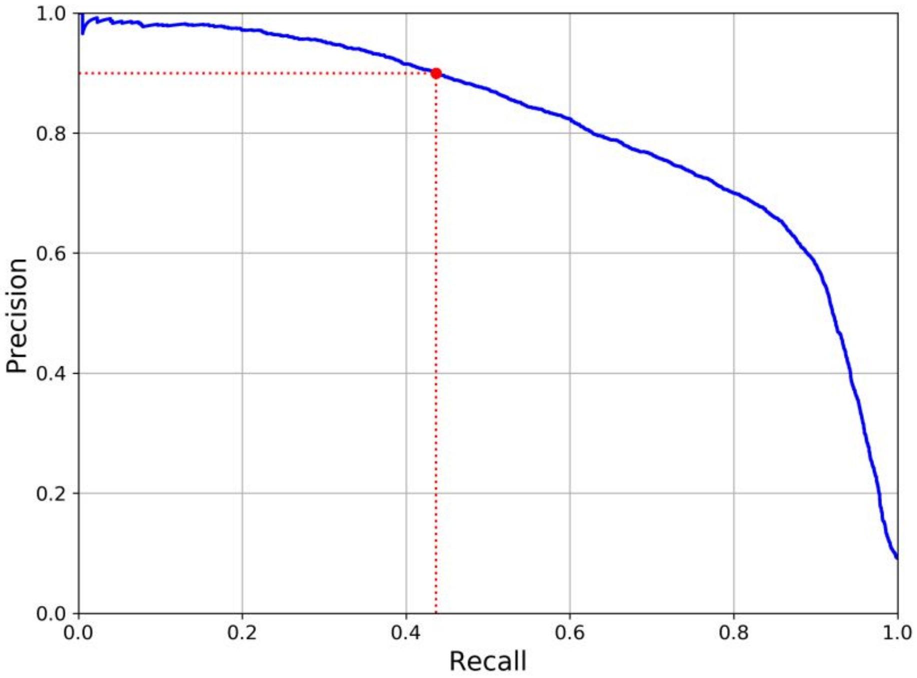
Precision/Recall Trade-off

How do you decide which threshold to use?



Precision/Recall Trade-off

How do you decide which threshold to use?



Precision/Recall Trade-off

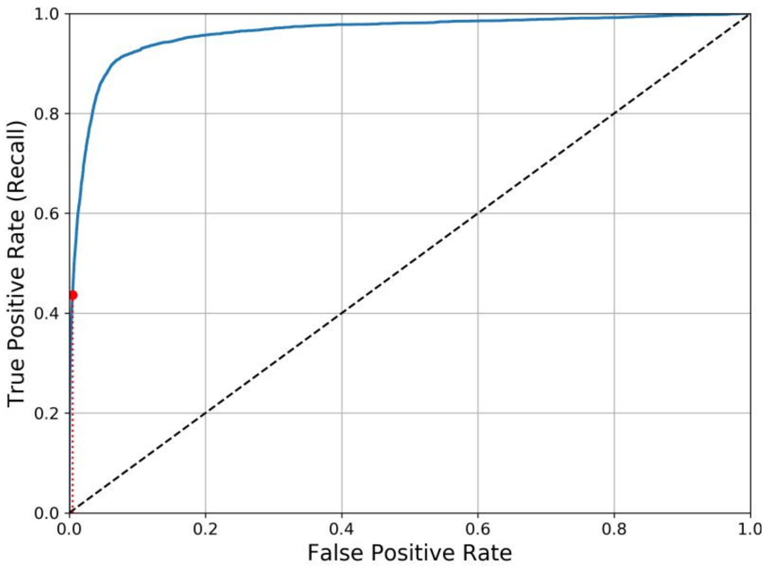
- A high-precision classifier is not very useful if its recall is too low!
- If someone says "let's reach 99% precision," you should ask, "at what recall?"

To take recall into consideration, we use other measures.

The ROC Curve

- The receiver operating characteristic (ROC) curve is another common tool used with binary classifiers.
- It is very similar to the precision/recall curve, but instead of plotting precision versus recall, the ROC curve plots the *true positive rate* (TPR, another name for *recall* or *sensitivity*) against the *false positive rate* (FPR, 1-*specificity*).

The ROC Curve



- Once again there is a tradeoff: the higher the recall (TPR), the more false positives (FPR) the classifier produces.
- The dotted line represents the ROC curve of a purely random classifier;
- A good classifier stays as far away from that line as possible (toward the top-left corner).

AUC: Area Under the (ROC) Curve

- One way to compare classifiers is to measure the area under the curve (AUC).
- A perfect classifier will have a ROC AUC equal to 1, whereas a purely *random classifier* will have a ROC AUC equal to 0.5.

Note: As a rule of thumb, you should prefer the PR (precision-recall) curve whenever the positive class is rare or when you care more about the false positives than the false negatives, and the ROC curve otherwise.

Next lecture

Loss Functions

22nd August 2023