Accuracy Measures (Evaluating a classifier's effectiveness)

Lesson 9

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Overview of Accuracy Measures

This topic involves ways for measuring how effective a classification model is

- Not as simple as counting how many predictions a classifier got right or wrong

Essential before deploying a model

Without accuracy measures, evaluating a model would be very subjective

- Essential for tweaking a model so that it is better for the data at hand

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Overfitting

A model can perform well in the training set but poorly on a test set. This is known as over-fitting

- Overfit models are not useful in any way
- It's easy to overfit models when having large datasets
- Accuracy measures can ensure that no overfitting takes place

CLASSIFICATION STATISTICS

Classification statistics are basically ways of analyzing the results of a classifier, in a methodical manner

They have to do with metrics, such as accuracy rate and error rate, that are commonly used

In general, they involve two main methods:

- –Confusion Matrix
- –ROC analysis (ROC curve)

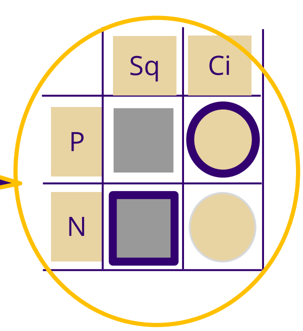
These methods are applicable for binary classification but can be leveraged with multi-class problems too, after some transformations

Confusion Matrix

CONFUSION MATRIX

What a CM looks like

Confusion Matrix
(Classification Matrix):
Vertical are actual classes
Horizontal are predicted classes



PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Constructing a CM

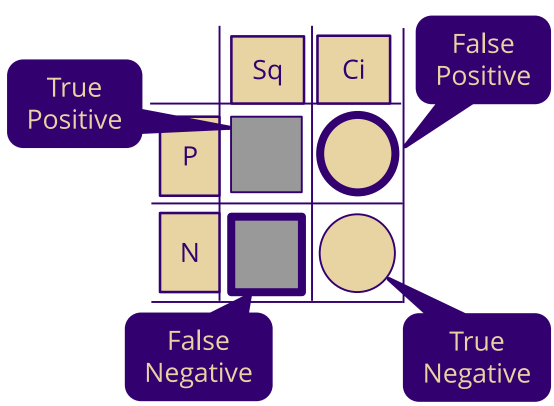
Organize a classifier's predictions in terms of how they relate with the two classes:

- 2 options for each class: correct or incorrect (depicted in chart as P (positive) and N (negative)
- Count how many predictions fall into each one of the 4 possibilities
- Name these counts accordingly

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

CONFUSION MATRIX

Frequency counts of 4 different possibilities



PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Metrics deriving from a CM

Value of CM: various metrics that derive from it, shedding light on different aspects of the classifier's performance

Which metric is best always depends on the problem at hand

Metrics deriving from a CM

Accuracy rate: all the correct predictions over the total predictions (sum of main diagonal elements over total elements).

– In other words: AR = (TP + TN) / N

Error rate: all the erroneous predictions over the total predictions.

– In other words: ER = (FP + FN) / N

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Metrics deriving from a CM

Precision: of all the predictions for a given class, what proportion of them the classifier got right.

– In other words: P = TP / (TP + FP). Reliability of predictions.

Recall: of all the elements related to a given class, what proportion of them the classifier got right.

– In other words: R = TP / (TP + FN). Net prediction potential.

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Metrics deriving from a CM

F1 score: an average of these two, leaning more towards the smaller one. Harmonic mean of P and R.

– In other words, F1 = 2*P*R / (P + R) = 2TP / (2TP + FP + FN)

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Metrics deriving from a CM

True Positive Rate: same as Recall. Aka “sensitivity”

False Positive Rate: equivalent to TP rate, but with FP.

– In other words, FPR = FP / (FP + TN).

TP rate and FP rate are negatively correlated to each other

> These metrics are useful in ROC analysis

> All of the CM metrics take values between 0 and 1 (inclusive) and are often expressed as percentages

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Optimizing for a given goal, based on a CM

You can create your own metric to optimize, for a particular problem

– E.g. if FP are more costly than FN, you can define a function like Z = 10FP + FN. Then you can use this function as your classifier's performance metric and try to minimize that.

> Important: always take into account both FP and FN in a function, otherwise you'll end up with a trivial classifier

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Relationship with Contingency Table (Crosstab)

Contingency Table (CT) = more generic version of CM, with multi-value variables represented in it

- A CT can include relative frequencies in it, instead of counts
- CTs are studied thoroughly in Statistics

In multi-class classification problems you can use a CT instead of a CM

- The accuracy measures need to be changed accordingly if you use a CT instead of CM

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Summary

> Always evaluate a model for Overfit
  - Works better on the training set than on the test set
> Confusion Matrix yields other metrics
  - Accuracy rate & error rate
  - Precision, recall, & F1
  - True positive rate & false positive rate
> Optimize the model to control both the FP and FN