Measuring Out Models

Metrics for Binary Classification
Binary Classification is arguably the most common and conceptually simple application fo machine learning in practice. However, there are still anumber of caveats in evaluating this simple task.

We are looking for the positive class.
We are not looking for the negative class.

Accuracy is not a good measure of predictive performance, since the number of mistakes one makes does not contain all the information that interests the data scientist.

Think of testing flight data for lateness. If the test is negative, then the observation is labelled "on time". But our predictive model might make mistakes…

It could classify a "delayed" flight as "on time." This incorrect positive prediction ("false positive") leads to some business costs involving travel. They could be as expensive as a few days stay in a hotel. Or it could have business consequences such as missing a timely opportunity.

The other possible mistake is that an "on time" flight could be classified as "delayed." This incorrect negative prediction is called a "false negative."

On the other hand, given similar flight data, we could also generalize a model that would predict airline catastrophes… But that is out of the realm of this presentation.

Reasons for using metrics:
Using a metric to assess a model ensures that it weeds out nonsense predictions.
Can easily spot balance or imbalanced classes.

Accuracy metrics for Logistic Regression:
Precision, Recall, f-score

Conclusions
Knowing the causes that contribute to airline delays creates better-informed consumers. Knowing specifically what types of delays are possible, when to expect delays, and what airlines and airports to associate with delays would be great information to have. This knowledge could empower the cutomer to choose between a list of the best performing airlines and airports given a set of conditions. Alternatively, the same customer could decide when to fly, expecially if his home airport is computed to be one of the main culprits. This knowledge could also empower the airlines so that they may monitor hencefoth or may research specific causes of

delays. Another business benefit could be saving time and capital expenditure (preventing loss). Another result could be attracting customers (increasing profit).

Because there is no association with life or death, we can say that it is not abundantly clear which consequence is more drastic, missed opportunites, travel expenditure over budget, profit/loss, etc. It all depends on the business decisions that have to be made based on the data. As such, we would have to assign dollar value to each kind of mistake, which would allow measuring data in actual money instead of abstract probability.