# Measuring performance in regression models

Classification models usually generate two types of predictions. Like regression models, classification models produce a **continuous valued prediction, which is usually in the form of a probability** (i.e., the predicted values of class membership for any individual sample are between 0 and 1 and sum to 1). In addition to a continuous prediction, classification models generate **a predicted class**, which comes in the form of a discrete category.

We desire that the **estimated class probabilities are reflective of the true underlying probability of the sample**. That is, the predicted class probability (or probability-like value) needs to be **well-calibrated**. To be well calibrated, the probabilities must effectively reflect the true likelihood of the event of interest.

One way to assess the quality of the class probabilities is using a **calibration plot**. For a given set of data, this plot shows some measure of the observed probability of an event versus the predicted class probability.

## Comparison of predictive ability:

An alternative approach is to test whether adding the new variable improves some measure of predictive ability, such as [the area under the ROC curve](http://thestatsgeek.com/2014/05/05/area-under-the-roc-curve-assessing-discrimination-in-logistic-regression/).

A common method for describing the performance of a classification model is **the confusion matrix**. This is a simple cross-tabulation of the observed and predicted classes for the data.

First, different important of the different classes. Second, one must consider the natural frequencies of each class.

The **no-information rate** is the accuracy rate that can be achieved without a model.

# logistic regression

# ROC curve

A **receiver operating characteristic** (**ROC**), or **ROC curve**, is a [graphical plot](https://en.wikipedia.org/wiki/Graph_of_a_function) that illustrates the performance of a [binary classifier](https://en.wikipedia.org/wiki/Binary_classifier) system as its discrimination threshold is varied. The curve is created by plotting the [true positive](https://en.wikipedia.org/wiki/True_positive) rate against the [false positive](https://en.wikipedia.org/wiki/False_positive) rate at various threshold settings.

**Sensitivity** and **specificity** are statistical measures of the performance of a [binary classification](https://en.wikipedia.org/wiki/Binary_classification) [test](https://en.wikipedia.org/wiki/Classification_rule), also known in statistics as [classification function](https://en.wikipedia.org/wiki/Statistical_classification). **Sensitivity** (also called the **true positive rate**, or the [**recall rate**](https://en.wikipedia.org/wiki/Precision_and_recall#Definition_.28classification_context.29) in some fields) measures the proportion of actual positives which are correctly identified as such (e.g., the percentage of sick people who are correctly identified as having the condition), and is [complementary](https://en.wikipedia.org/wiki/Complementary_event) to the [false negative rate](https://en.wikipedia.org/wiki/False_negative_rate). **Specificity** (sometimes called the **true negative rate**) measures the proportion of negatives which are correctly identified as such (e.g., the percentage of healthy people who are correctly identified as not having the condition), and is complementary to the [false positive rate](https://en.wikipedia.org/wiki/False_positive_rate).

Let us define an experiment from **P** positive instances and **N** negative instances for some condition. The four outcomes can be formulated in a 2×2 [*contingency table*](https://en.wikipedia.org/wiki/Contingency_table) or [*confusion matrix*](https://en.wikipedia.org/wiki/Confusion_matrix), as follows: