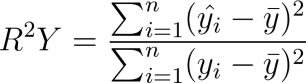
# OPLS-DA outputs

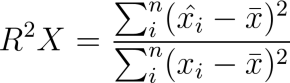
## R2

### Description:

* R2 the coefficient of multiple determination, is a measure of the adequacy of a fitted model.
* R2X is the amount of X variation in the X data (variables / bins / metabolites) explained by the model. This X variation is independent of the classes and is likely to be noise.
* R2Y is the amount of Y variation in X data (classes / treatment vs control) explained by the model. This is what we want to capture in the model. It is the information most vital in distinguishing between the classes.

### Formula:





* y\_hat and x\_hat are the predicted values for y and x respectively
* y\_bar and x\_bar are the mean values for y and x respectively
* y and x are the actual y and x values respectively

### Guidelines for good values:

Ideally, we'd like R^2X to be low, closer to zero. This would indicate that the OPLS model has excluded some amount of X variation that does not pertain to the class information.

We'd like R^2Y to be closer to 1.  This would indicate that a large amount of the Y variation, pertaining directly to the differences between classes, has been explained by the model.

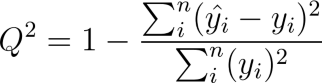
## Q2

### Description:

Q^2 is a cross-validation metric which characterizes the predictive power of a model.  It is calculated using a leave-one-out cross validation.

Leave-one-out cross-validation is performed as follows.  Say you have 10 treatment samples and 10 control samples.  One of the samples is set aside and the model is created (using OPLS-DA) using the remaining 19 samples.  The left out sample's class is predicted using this model and how close the predicted sample its actual class is recorded.  This sample is then put back in the group and a new sample is removed.  The process is repeated until each sample has been left out once.

### Formula:



### Guidelines for good values:

A Q^2 of less than 0 indicates the model has no predictive power.  A value above 0 indicates the model has some predictive power.  The closer to 1, the more predictive power and the more reliability the model has.

## Q^2 from permutation of class labels

### Description:

This permutation test is designed to test for overfitting.  The class labels are randomly assigned (keeping consistent the number of treated and control samples) and a leave-one-out Q^2 is computed.  This is repeated some number of times.  We then have a distribution of permuted Q^2 values.  A confidence interval is calculated and the correctly assigned Q^2 is compared to this CI.

### Guidelines for good values:

The correctly assigned Q^2 should be outside of the CI.  This indicates that the model is not overfit.

## Accuracy

### Description:

Accuracy is the percentage of correctly identified samples during cross validation.

Accuracy is determined using leave-one-out cross-validation as described above.  Here though, we are only concerned with whether the model predicted the class correctly or not, whereas before we were concerned with how far off the predicted was from the actual even if it predicted the absolute class correctly.

## Plots

### X scores ( T ) versus observation

Demonstrates seperation, if any is determined by OPLS-DA, between the classes.  Dividing line is likely to be at 0 on the Y-axis.

### Orthogonal X scores ( T orthogonal ) versus X scores ( T )

Figure demonstrates the separation on the data filtered by OPLS-DA versus the data deemed important by OPLS-DA.  Along the X-axis, we would like to see separation of the classes, and in fact should see a similar separation as in the X scores versus observation plot.  Along the Y-axis, we don't want to see separation.  The Y-axis represents data that was filtered out because it did not capture any Y variation (it did not help to separate the classes).  The data from the two classes should overlap by a large amount on this axis.

### X loading ( P ) versus chemical shift

Indicates the bins with the largest loadings, i.e. the most significant bins.  Loadings plotted versus bin position to imitate the look of a spectrum. Absolute value is all that is important here.  Positive and negative signs do not really give any extra insight.

# Sample write-ups



