R Functions in Session 2

Business Analytics (S19)

cor(x, y): Compute the correlation of vector x and vector y.

var(x, y, na.rm = FALSE): Compute the variance of vector x and vector y.

cov(x, y): Compute the covariance of vector x and vector y.

set.seed(x): Sets the seed of R‘s random number generator, which is useful for creating simulations or random objects that can be reproduced. Please try to understand it by reading codes.

example:

```
> rnorm(5) # don’t set seed, objects are different.
[1] -0.93415132  1.32360565  0.62491779 -0.04572296 -1.00412058
> rnorm(5)
[1] -0.8284332 -0.3483517 -1.5382934 -0.2555652 -1.1499450

> set.seed(11) # set seed, objects are same. “11” just a id
> rnorm(5)
[1] -0.59103110  0.02659437 -1.51655310 -1.36265335  1.17848916
> set.seed(11)
> rnorm(5)
[1] -0.59103110  0.02659437 -1.51655310 -1.36265335  1.17848916
```

sample.split(Y, SplitRatio = 2/3, group = NULL)： Splits data from vector Y into two sets in predefined ratio while preserving relative ratios of different labels in Y. Used to split the data used during classification into train and test subsets.

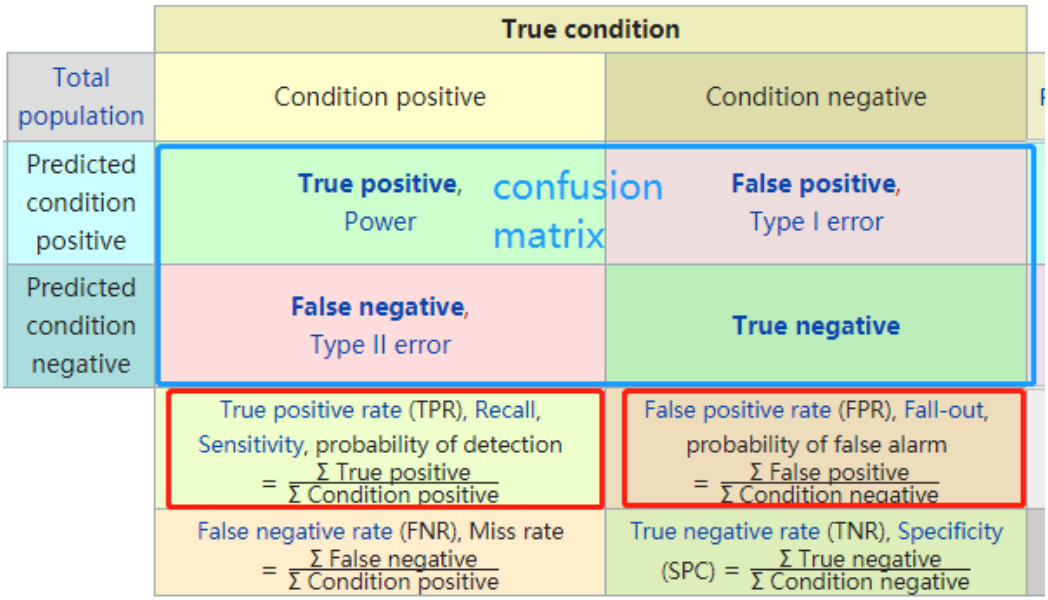
glm(formula, family =binomial, data, control = list(...): Fits generalized linear models, specified by giving a symbolic description of the linear predictor and a description of the error distribution.

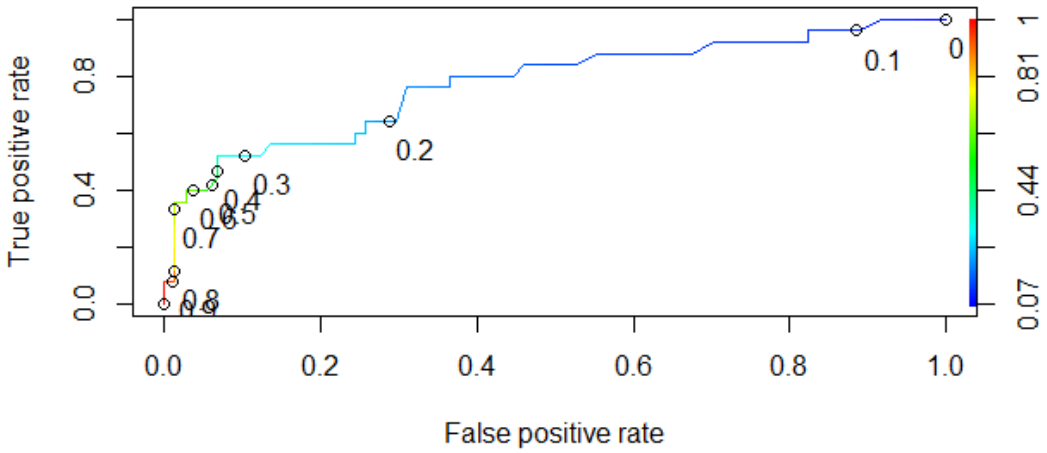
family: a description of the error distribution and link function to be used in the model. Here are three parameters could be chosen, binomal, Gamma, poisson. In this session, you should choose binomal, because binomal represents logistic regression.

prediction(predictions, labels, label.ordering = NULL): Every classifier evaluation using ROCR starts with creating a prediction object. This function is used to transform the input data (which can be in vector, matrix, data frame, or list form) into a standardized format.

performance(prediction.obj, x.measure,y.measure): All kinds of predictor evaluations are performed using this function.

After using performance function, we often use plot(performance object) to plot a ROC curve. Please try to understand the following pictures, especially the definition of the TPR and FPR. Then you could plot a ROC curve.





as.numeric(objects): Creates or coerces objects of type "numeric"

as.factor(): Encodes a vector as a factor (the terms ‘category’ and ‘enumerated type’ are also used for factors). If argument ordered is TRUE, the factor levels are assumed to be ordered.

roc(predictions, labels): Computes the receiver operating characteristic (ROC) curve required for the aucfunction and the plot function.

auc(x): Computes the area under the sensitivity curve (AUSEC), the area under the specificity curve (AUSPC), the area under the accuracy curve (AUACC), or the area under the receiver operating characteristic curve (AUROC).