**Key Points**

A1:

**1. Multicollinearity** refers to a situation in which two or more explanatory variables in a multiple regression model are highly linearly related. We have perfect multicollinearity if, for example as in the equation above, the correlation between two independent variables is equal to 1 or −1. In practice, we rarely face perfect multicollinearity in a data set. More commonly, the issue of multicollinearity arises when there is an approximate linear relationship among two or more independent variables.

Variance inflation factor (VIF) is used to detect multicollinearity.

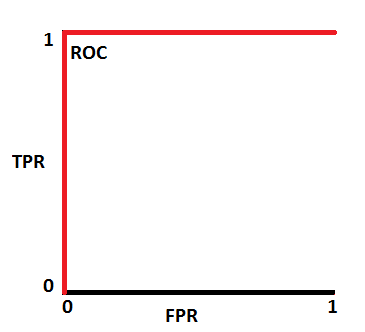
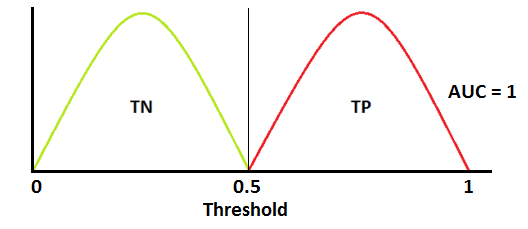
Where is the coefficient of determination of a regression of explanator on all the other explanators. A VIF of 5 or 10 and above indicates a multicollinearity problem.

Ref: <https://en.wikipedia.org/wiki/Multicollinearity>

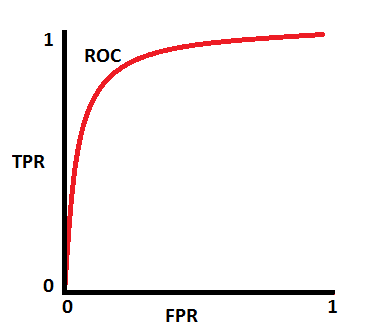
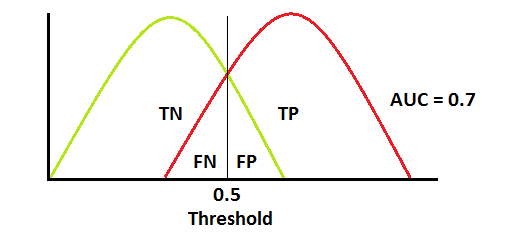
**2. AUC-ROC curve** is a performance measurement for classification problem at various thresholds settings. ROC is a probability curve and AUC represents degree or measure of separability. It tells how much model is capable of distinguishing between classes. Higher the AUC, better the model is.

As we know, ROC is a curve of probability. So lets plot the distributions of those probabilities:

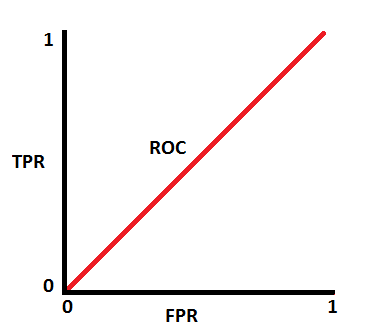
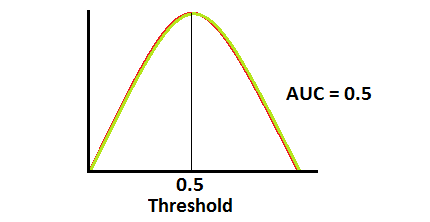
Note: Red distribution curve is of the positive class (patients with disease) and green distribution curve is of negative class(patients with no disease).



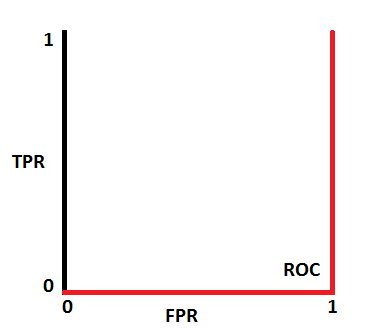
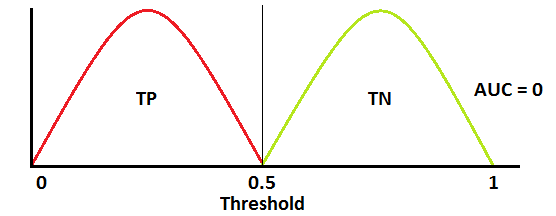
This is an ideal situation. When two curves don’t overlap at all means model has an ideal measure of separability. It is perfectly able to distinguish between positive class and negative class.



When two distributions overlap, we introduce type 1 and type 2 error. Depending upon the threshold, we can minimize or maximize them. When AUC is 0.7, it means there is 70% chance that model will be able to distinguish between positive class and negative class.



This is the worst situation. When AUC is approximately 0.5, model has no discrimination capacity to distinguish between positive class and negative class.



When AUC is approximately 0, model is actually reciprocating the classes. It means, model is predicting negative class as a positive class and vice versa.

Ref: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>

TBC