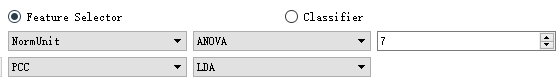
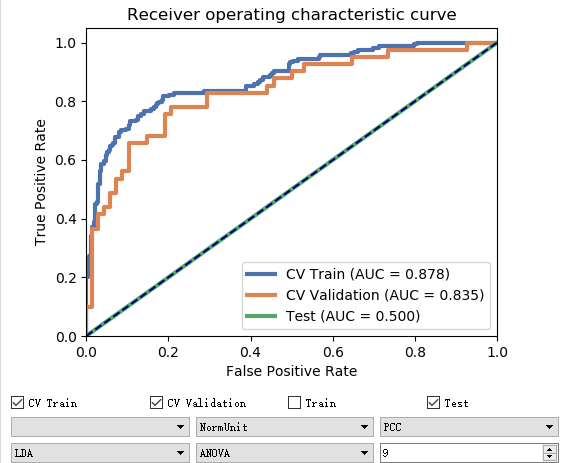
1. Fae中classifier和feature slector的区别是什么？

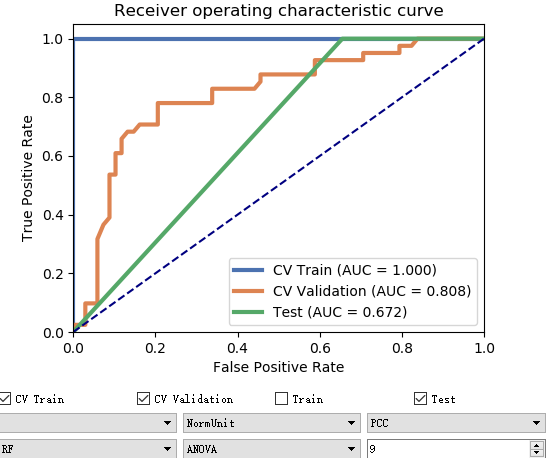
我之前一直认为feature selecor会筛选出分类性能从高到低的feature，然后模型会依次选取前几名的特征进行训练。但今天才发现有写模型似乎最后选取的是classifier中的特征。如果这样的话，二者的区别与另一方存在的意义是什么。



1. 一些模型会出现在test集中, 预测值十分集中的情况。

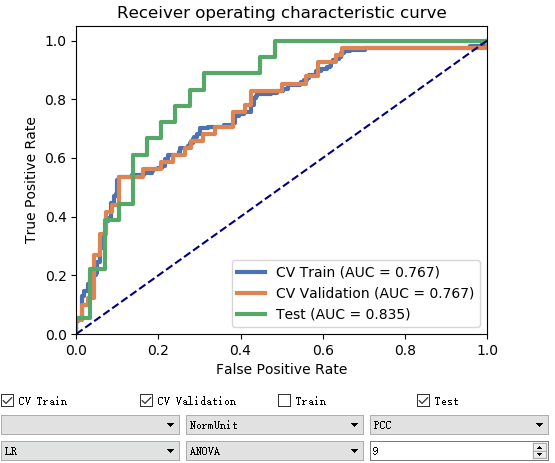
如下图，roc曲线的折点非常少。甚至LDA模型中所有的test都预测成了1，但是训练集和测试机中都没有什么问题。另外有的模型会将test的预测值只会在几个小数之间取值。我不太想的明白，为何在train、validation正常的情况下，test集会集中预测在几个值上。





1. Test表现比train 和 validation好。

如下图，逻辑回归模型中，无论feature number取几，test的表现都跟好，个人十分不理解。



1. 这次训练出来的模型中，validation结果最好的模型前20名几乎被PCA刷屏，但是test上反过来被PCC刷屏，为何PCA会造成如此严重的过拟。

关于这一点我心中的有一点点眉目，毕竟PCA多经过了一次降维处理，自然更容易过拟，但应该也达不到如此严重的地步。不知道我的这个想法对不对，或者其他的说法更加合理。