Exercise 3:

通过上面方法的 ROC 图我们观察到 Ensemble Methods 有最好的分类结果。

（1） 我们选取了 Decision Trees、AdaBoost、RandomForest 方法并画出了各个方法排名前五的特征，如上图所示。

（2） 土地覆被测绘是地球观测卫星传感器的主要应用之一，它利用遥感和地理空间数据来识别位于目标区域表面的材料和物体。通常，目标材料的类别包括道路，建筑物，河流，湖泊和植被。基于人工神经网络的一些不同的集成学习方法，，带 Boosting 的决策树，随机森林和多分类器系统的自动设计，被提出来有效地识别土地覆盖物。

（3） 随机森林可以处理很高维度的数据，并且不用降维，无需做特征选择，且可以判断特征的重要程度，训练速度较快

（4） 随机森林算法在某些噪音较大的分类或回归上会过拟合，对于不同取值属性的数据，取值划分较多的属性会对随机森林产生更大的影响。

（5） 在这个数据集中，数据的维度比较高，所以使用随机森林算法可以比较快速的完成拟合，在随机森林算法处理完可以从中得出重要特征，更利于分析这个问题得出收入高低的主要影响因素。