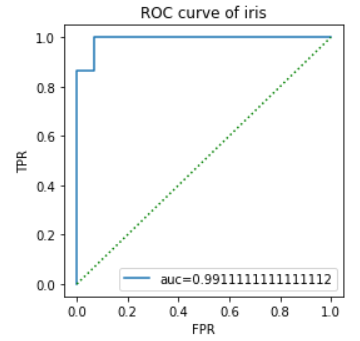
**Iris数据曲线**

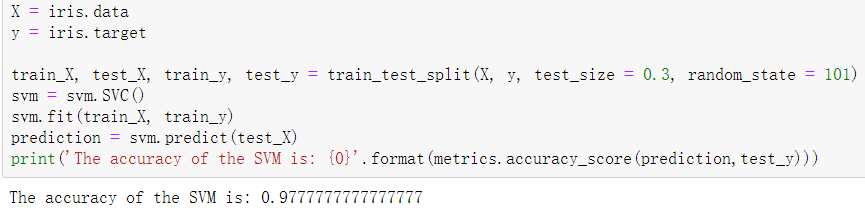
1. ROC曲线

（1）二分类

选取标签为1和2的数据集画曲线，利用支持向量机分类器，找到数据到超平面的距离，再利用roc\_curve计算得出阈值，根据阈值的变化画出曲线。由于数据量较小故曲线不平滑，但最后该数据集得到的AUC为0.99，曲线也接近于左上角，故该分类器效果较好。

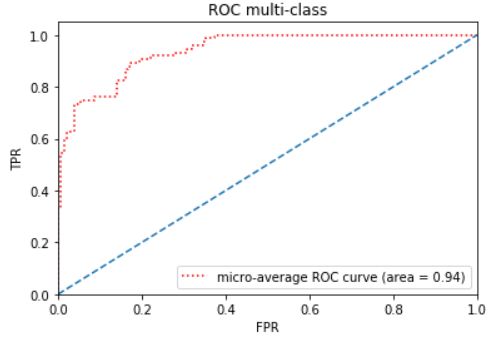


支持向量机分类器对于整个数据准确率也很高



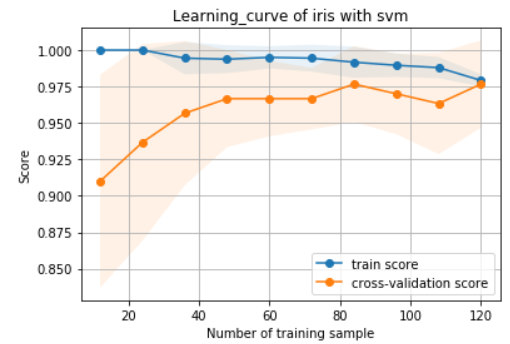
（二）多分类

将每个样本都作为二分类计算



1. 学习曲线（偏差方差）

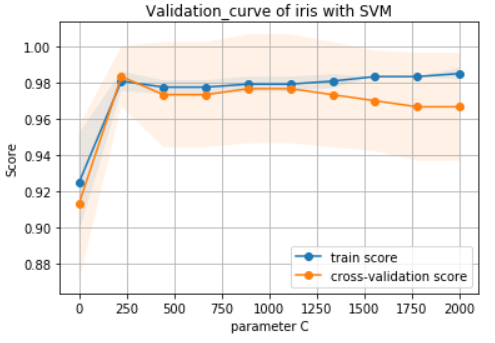
选取模型为支持向量机分类器，数据间进行交叉验证，随着训练样本的增加，训练曲线和验证曲线收敛于0.975，由于训练分数一直较高，所以可能会有些过拟合。



1. 验证曲线（欠拟合过拟合）

选择支持向量机作为模型，依旧是交叉验证。由于要选择一个超参数作为横轴自变量，故选择SVC里的惩罚参数C，根据param\_range的调整选择最合适的C值。

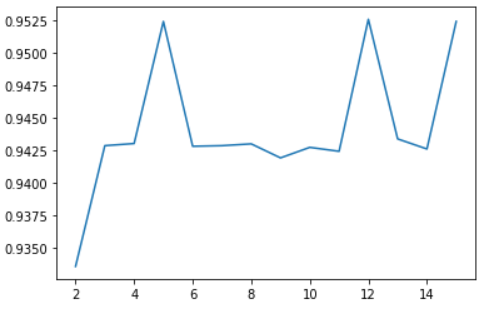
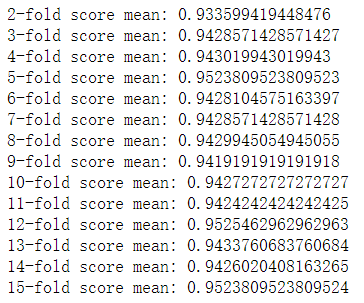
最初训练集和交叉验证准确率都较低，为欠拟合，到达某个点后，训练集准确率上升，而叫交叉验证准确性开始下降，此为过拟合。最后选择一个变化的转折点作为C的最优选择。该模型选择C=1100。



1. K-fold交叉验证和参数寻优

（1）KFold

以支持向量机作为模型，利用cross\_val\_score方法计算出k为2-15的平均分数，并画图折线图，当k=5、12、15时分数都较高。



（2）参数寻优

利用GridSearchCV方法对指定的超参数列表进行穷举搜索，评估每个组合对模型性能的影响，最终得到svm模型在超参数svc\_\_C=10.0的时候可得到最优k折交叉验证准确率96.7%。最后用最优参数进行模型性能评估得到准确率为100%。

