数据处理说明：

SVM算法直接处理的数据格式是LIBSVM类型，要把原txt文件中的文字信息进行格式的转换。

1. 对于训练集和测试集都进行特征值的量化，age、fnlwgt等取值连续的特征对应的特征值保持不变，workclass、education等取值离散且用文字代表的特征进行量化，用不同的数字代表不同的特征值。类似的，训练集中最后一列的分类信息也用0和1代表收入在50k上下两类。将数据集变为全数字的txt文件。
2. 将以上得到的txt文件转换为LIBSVM格式。借用FormatDataLibsvm.xls来实现。首先通过excel打开txt文件，把文件中的分隔符（逗号）去掉，把文件分隔成一列列分离独立的数据，同时把txt文件通过excel转换成.xls类型的文件。接着打开.xls类型的文件，运行FormatDatatoLibsvm的宏把数据转为LIBSVM类型，并另存为txt文件。这样就变为MLlib中SVM算法可直接处理的数据类型。

运行结果中0代表<=50k，1代表>50k。