Work II

小组成员：冯宇婷 2111712336 专硕1701

黄金星 2111712360 专硕1702

施佳琴 2111712341专硕1702

张雷雷 2111712343 专硕1702

**肌电信号实验报告**

肌电图（electromyography）（EMG），应用电子学仪器记录肌肉静止或收缩时的电活动，及应用电刺激检查神经、肌肉兴奋及传导功能的方法。英文简称EMG。通过此检查可以确定周围神经、神经元、神经肌肉接头及肌肉本身的功能状态。本实验的目的是对采集到的肌电信号进行分类，主要包括采集数据，形成数据集和测试集，以及模型训练三部分。

首先对于数据采集部分，精确的动作产生的信号更加便于分类，我们规定采集的动作方式，正式采集开始前，按照选定的采集动作数量设置保存时间，假如需要采集10组动作，则设置为100000（100s），假如采集8组动作，则设置80000（80s）。每组动作占据10s时长。这里我们做了七组动作：

0-5s：休息，手部保持放松状态

6-10s：动作一

11-15s：休息，手部保持放松状态

16-20s：动作二

21-25s：休息，手部保持放松状态

26-30s：动作三

31-35s：休息，手部保持放松状态

36-40s：动作四

41-45s：休息，手部保持放松状态

46-50s：动作五

51-55s：休息，手部保持放松状态

56-50s：动作六

61-65s：休息，手部保持放松状态

66-60s：动作七

其次，读取采集的信号并按7:3的比例形成训练集和测试集，代码如图1

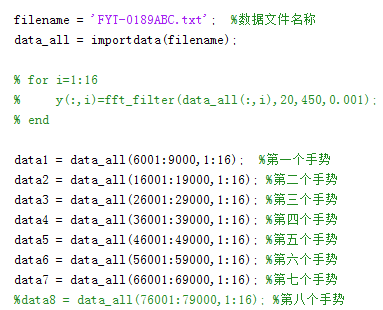


图1-a 数据读取

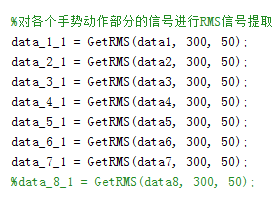


图1-b 提取RMS信号

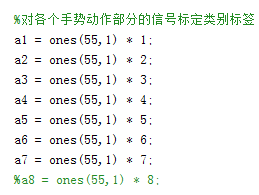


图1-c 标定标签

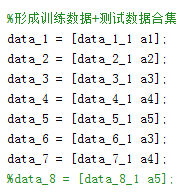


图1-d 形成训练数据+测试训练合集

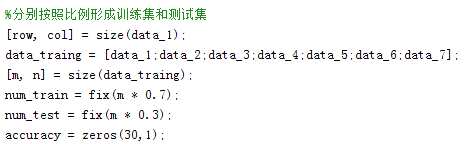


图1-e 形成训练集和测试集

其中RMS信号提取部分代码如图2

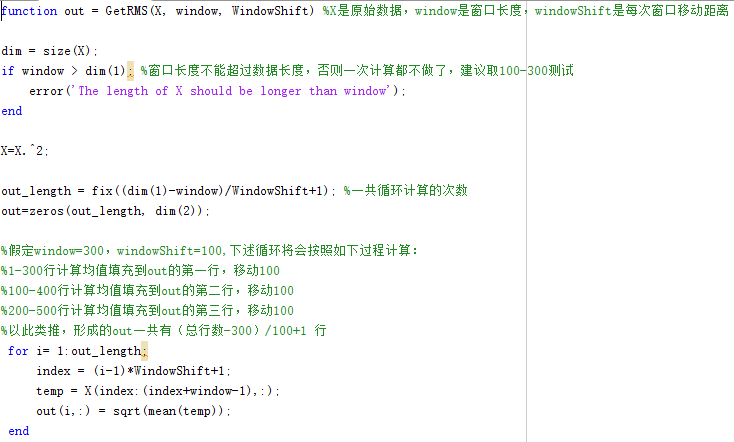


图2 RMS信号的提取

最后，通过实验对比，我们发现LDA的分类模型效果最好，其分类模型代码如图3，其验证代码如图4，验证结果如图5

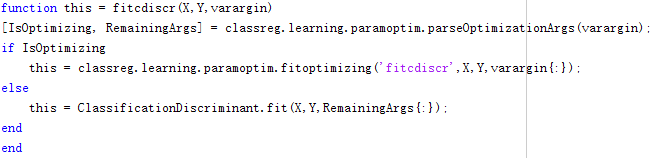


图3 训练模型代码

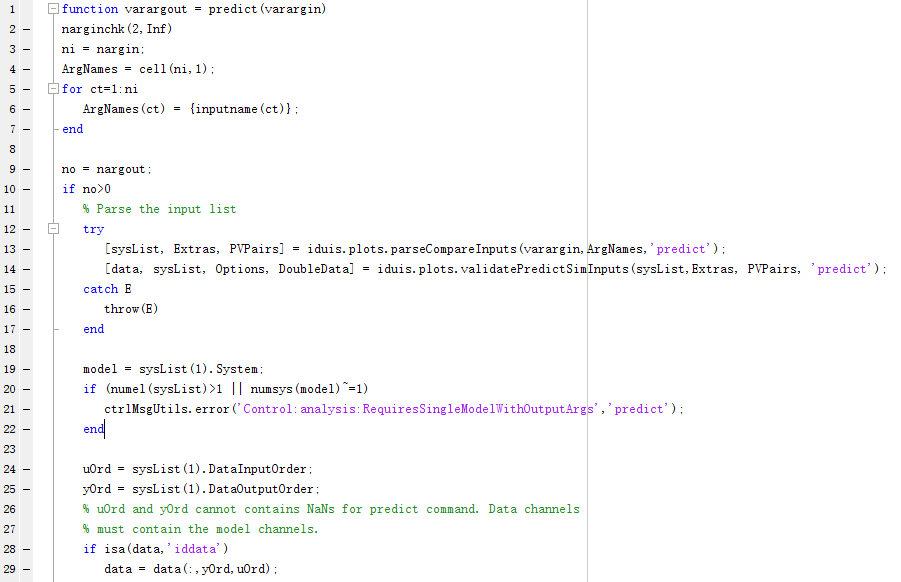


图4-a 测试模型代码

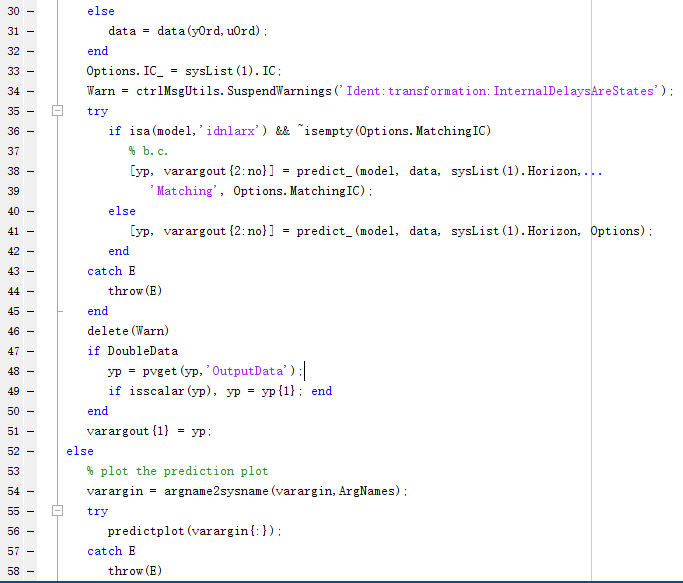


图4-b测试模型代码

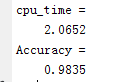


图5 测试结果

通过采用不同特征的提取方法和分类方法进行实验，其中包括，LDA分类，SVM分类，knn分类，bayes分类，我们发现各个分类方法都可以达到较高的精度，且各有优点，但对于本实验的肌电信号分类中，LDA的效果相对更好,以下是我们对LDA（线性鉴别分析的理解）。

LDA(性鉴别分析)采用的是有监督的方式，它先对训练数据进行降维，然后找出一个线性判别函数，其基本思想是将高维的模式样本投影到最佳鉴别矢量空间，以达到抽取分类信息和压缩特征空间维数的效果，投影后保证模式样本在新的子空间有最大的类间距离和最小的类内距离，即模型在该空间中有最佳的可分离性。因此，它是一种有效的特征抽取方法。使用这种方法能够使投影后模型样本的类间散度矩阵最大，并且同时类内散度矩阵最小。就是说，它能够保证投影后模型样本在新的空间中有最小的类内距离和最大的类间距离，即模式在该空间中有最佳的可分离性。

首先给定特征为d维的N个样例，，其中有个样例属于类别，另外个样例属于类别。现在我们觉得原始特征数太多，想将d维特征降到只有一维，而又要保证类别能够“清晰”地反映在低维数据上，也就是这一维就能决定每个样例的类别。假设这个最佳映射向量为w（d维），那么样例x（d维）到w上的投影可以用式（1）表示

（1）

假设样本向量包含2个特征值（d=2），我们就是要找一条直线（方向为w)来做投影，然后寻找最能使样本点分离的直线。

接下来要从定量的角度来找到这个最佳的w。

首先，每类样本的投影前后的均值点分别为(此处样本总数C=2),Ni表示每类样本的个数：

（2）

（3）

由式（2）（3）可知，投影后的的均值也就是样本中心点的投影。

其次，能使投影后的两类样本均值点尽量间隔较远的就可能是最佳的投影向量，将其定量表示就是：

（4）

此外，还需加一个约束—散列值，即对投影后的类求散列值，如式（5）：

（5）

从公式中可以看出，只是少除以样本数量的方差值，散列值的几何意义是样本点的密集程度，值越大，越分散，反之，越集中。而我们想要的投影后的样本点的样子是：不同类别的样本点越分开越好，同类的越聚集越好，也就是均值点间距离越大越好，散列值越小越好。正好，我们可以使用J(w)和S(w)来度量。因此可以定义最终的度量公式如式（6）：

（6）

我们只需寻找使最大的w即可。

但这种方法也存在一定的缺陷，比如它不适合对非高斯分布的样本进行降维，且在样本分类信息依赖方差而不是均值时效果不够好，当然也可能会过度拟合数据。