

第二次作业（在网上提交电子版）

（本次作业覆盖第 2 章和第 3 章的内容）

第一部分：简述题（必做）

1. 请描述对数几率回归模型（包含学习任务、学习模型、假设空间）。
2. 请描述正则化 softmax 回归模型（包含学习任务、学习模型、假设空间）。
3. 请描述 ECOC 的优缺点。
4. 请描述软间隔支持向量机的原始模型及其对偶模型，并指出满足哪些条件的样本是支持向量。
5. 请简述核支持向量机的原理。

第二部分：计算机编程（采用 Matlab）：（第 1 题和第 2 题选做一道）

第 1 题：

首先，使用 Matlab 生成如下数据集：

```
Sigma = [1, 0; 0, 1];  
mu1 = [1, -1];  
x1 = mvnrnd(mu1, Sigma, 200);  
mu2 = [5, -4];  
x2 = mvnrnd(mu2, Sigma, 200);  
mu3 = [1, 4];  
x3 = mvnrnd(mu3, Sigma, 200);  
mu4 = [6, 4.5];  
x4 = mvnrnd(mu4, Sigma, 200);  
mu5 = [7.5, 0.0];  
x5 = mvnrnd(mu5, Sigma, 200);  
  
% Show the data points  
plot(x1(:,1), x1(:,2), 'r.'); hold on;  
plot(x2(:,1), x2(:,2), 'b.');  
plot(x3(:,1), x3(:,2), 'k.');  
plot(x4(:,1), x4(:,2), 'g.');  
plot(x5(:,1), x5(:,2), 'm.');
```

现在假定 x_1 中的 200 个点属于第一类， x_2 中的 200 个点属于第二类， x_3 中的 200 个点属于第三类， x_4 中的 200 个点属于第四类， x_5 中的 200 个点属于第五类。请编写一个 Matlab 程

序，采用正则化 softmax 回归方法来实现分类。

要求：

- (1) 每类随机采用 60%的数据做训练样本，采用剩余的 40%样本作测试。请给出正确率随着正则化参数而改变的曲线或者表格。
- (2) 采用交叉验证来选择正则化参数（采用 matlab 实现）。

注：请将代码同时拷贝到作业纸上提交。

第二题：

对第一题中的数据，请采用 SVM 对其进行分类（会调用 LibSVM 即可）。

要求：

- (1) 每类随机采用 60%的数据做训练样本，采用剩余的 40%样本作测试。请给出正确率随着正则化参数而改变的曲线或者表格。
- (2) 采用交叉验证来选择正则化参数（采用 matlab 实现）。