

1、spamTrain.mat 是一组关于垃圾邮件的训练数据，spamTest.mat 对应的则是测试数据。（当前数据中的标签为 0 和 1，为了与授课内容保持一致，可以将训练数据和测试数据的标签调整为-1 和 1）。令 $C=\{0.1, 1, 10\}$ ，试使用软-SVM 的高效求解算法佩加索斯（Pegasos）算法训练一个分类器，判断测试数据中的每个样本是否是垃圾邮件，并计算预测精度。（10 分）

2、试根据授课内容，使用坐标下降法给出 Lasso 问题： $\min_{\beta, \beta_0} \sum_{i=1}^n (y_i - \beta^T x_i - \beta_0)^2 + \lambda \|\beta\|_1$ 求解的详细推导步骤（5 分）。

3、orl_faces 数据集包含四十个文件夹，每一个文件夹中有 10 个人脸图像。试根据字典学习问题：

$$\min_{B, \alpha_i} \sum_{i=1}^n (\|x_i - B\alpha_i\|^2 + \lambda \|\alpha_i\|_1)$$

学习人脸图像的稀疏表示。在上式中 x_i 表示由 *.pgm 读出的人脸样本原始表示。 $B \in R^{p \times k}$ 为字典矩阵， k 为字典的词汇量， $\alpha_i \in R^k$ 是样本 $x_i \in R^p$ 的稀疏表示。其中，字典 B 的第 i 列通过第 i 个文件夹中的人脸随机初始化（ $k = 100$ ）， $\lambda = \{0.1, 1, 10\}$ 。（10 分）