

# 实验课题 1--支持向量机 SVM

## 一、课题目的

1. 理解 SVM 原理，掌握基于 Hinge Loss 的 SVM 损失函数（也称为代价函数）的推导过程，并能推导损失函数的梯度。
2. 掌握基于梯度下降法的 SVM 优化方法。
3. 能实现在 Breast Cancer 数据集上的分类任务，掌握数据预处理处理方法。

## 二、SVM 理论

参考课程 PPT 资料以及实验 PPT 资料。

## 三、课题内容

### 1. 给定数据集 Breast Cancer，该数据集的数据格式如下：

数据集：breast-cancer\_scale.txt （特征已归一化为[-1,1]区间）

资料来源：UCI / Wisconsin Breast Cancer 数据

#类数： 2

#数据样本数： 683

#特征维度： 10

类别标签=2，表示癌症

类别标签=4，表示不是癌症

数据集文件中的每一行数据：

```
classLabel featureID1: featureValue1 featureID2: featureValue2 ... featureID10: featureValue10
```

### 2. 课题要求

（1）利用给定的部分代码程序（由 svm\_start.py 提供），补充完成该程序（在#write your code...位置补充完成程序），实现一个支持向量机(SVM)分类器，用于分类癌症(标签为 4)或非癌症(标签为 2)，需要使用梯度下降来训练 SVM。

（2）数据使用：请使用前 400 个（行）数据用于训练模型，剩下的 283 个数据用于测试模型性能。

（3）测试完成训练的模型性能，利用准确率指标。

（4）编程语言与环境：Python，Anaconda，需要安装 matplotlib 模块。

**注意 1:**其他编程语言是可以接受的。

**注意 2:**不允许使用机器学习库(如 scikit-learn)进行训练和测试。

(4) 撰写实验课题报告，下面是实验课题报告的简单模版。

---

	题目	
班级:	学号:	姓名:

一、课题目的

二、SVM 算法理论

三、实验结果与分析

应包括模型准确率，不同的超参设置对 SVM 性能的影响，模型训练曲线等分析。

四、总结实验心得

请注意：除了实验课题报告，还需要提交的附件：[源代码（请提供算法关键步骤的语句解释）](#)。