# 机器学习第六章实验内容

详细内容见第五周实验内的jupyter notebook。

## 一、实验内容

### 1.1 支持向量机的核函数

1. 了解核函数对SVM的影响
2. 绘制不同核函数的决策函数图像
3. 简述引入核函数的目的

### 支持向量机的软间隔

1. 了解分离超平面、间隔超平面与支持向量的绘制
2. 调整C的值，绘制分离超平面、间隔超平面和支持向量
3. 简述引入软间隔的原因，以及C值对SVM的影响

### 1.3 支持向量机的分类任务

1. 使用支持向量机完成spambase垃圾邮件分类任务
2. 使用训练集训练模型，计算测试集的精度，查准率，查全率，F1值

### 1.4支持向量机的回归任务

1. 使用支持向量机完成kaggle房价预测问题
2. 使用训练集训练模型，计算测试集的MAE和RMSE

### 1.5完成dota2比赛结果预测任务（选做）

1. 使用你学过的算法完成dota2比赛结果预测
2. 对比它们的精度、查准率、查全率、F1值

**要求将结果写入到markdown的表格中！**

## 二、数据介绍

### 2.1 kaggle房价预测数据集

文件名：

1. 原始数据：kaggle\_hourse\_price\_train.csv
2. 字段说明：kaggle房价预测字段说明.txt

数据来源：<https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>

### 2.2 spambase垃圾邮件数据集

文件名：

1. 原始数据：spambase.data
2. 字段说明：spambase数据说明.txt

数据来源：<http://archive.ics.uci.edu/ml/datasets/spambase>

数据无需预处理，读取即可使用

import numpy as np

data = np.loadtxt('data/spambase/spambase.data', delimiter = ",")

### 2.3 Dota2 Games Results Data Set Dota2游戏结果数据集

文件名：

1. 原始数据：dota2Train.csv
2. 字段说明：dota2比赛结果字段说明.txt

数据来源：<http://archive.ics.uci.edu/ml/datasets/Dota2+Games+Results>

数据无需预处理，读取即可使用

import numpy as np

data = np.loadtxt('data/dota2Dataset/dota2Train.csv', delimiter=',')