

统计学习理论及应用课程项目介绍

一、项目说明

本课程有 4 个项目：(a) 回归模型项目；(b) 支持向量机项目；(c) 多层感知机项目；(d) Adaboost 项目。

二、提交时间与方式



三、具体要求

3.1 回归模型项目

项目要求：请采用回归模型，对“regression”目录下的数据进行建模。回归模型的复杂度与求解方法，需自己确定并编程实现。不能采用编程语言或第三方现有编程模型。

数据集有三个数据文件，均为 csv 格式，其中训练集有 10000 个样本，验证集和测试集均有 1000 个样本。训练集和验证集包含了标签信息，用于同学们训练和验证模型；测试数据集用于评价大家模型的学习性能。请同学提交训练结果时，将测试集的预测结果写入“regression_test.csv”文件的 y 列中，回归预测结果为实数。

编程语言：自选，建议采用 Matlab 或 Python。

项目评分：回归数据集采用均方误差，以及源文件的可读性作为评价指标。

提交内容：程序源文件，“regression_test.csv”文件。

3.2 支持向量机项目

项目要求：请采用支持向量机模型，对“classification”目录下的数据进行建模。支持向量机模型，需要自己确定。支持向量机模型可采用他人模型，模型的其它超参数，如惩罚因子 C，核函数等，需要自己确定。

数据集有三个数据文件，均为 csv 格式，其中训练集有 10000 个样本，验证集和测试集均有 1000 个样本。训练集和验证集包含了标签信息，用于同学们训练和验证模型；测试数据集用于评价大家模型的学习性能。请同学提交训练结果时，将测试集的预测结果写入“svm_classification_test.csv”文件的 y 列中，分类结果为 0 或 1。

编程语言：自选，建议采用 Matlab 或 Python。

项目评分：分类数据集采用正确率作为评价指标，以及源文件的可读性作为评价指标。

提交内容：程序源文件，“smv_classification_test.csv”文件。

3.3 多层感知机项目

项目要求：请采用多层感知机模型，对“classification”目录下的数据进行建模。多层感知机模型，需自己确定并编程实现。不能采用编程语言或第三方现有编程模型。

数据集有三个数据文件，均为 csv 格式，其中训练集有 10000 个样本，验证集和测试集均有 1000 个样本。训练集和验证集包含了标签信息，用于同学们训练和验证模型；测试数据集用于评价大家模型的学习性能。请同学提交训练结果时，将测试集的预测结果写入“mlp_classification_test.csv”文件的 y 列中，分类结果为 0 或 1。

编程语言：自选，建议采用 Matlab 或 Python。

项目评分：分类数据集采用正确率作为评价指标，以及源文件的可读性作为评价指标。

提交内容：程序源文件，“mlp_classification_test.csv”文件。

3.4 Adaboost 项目

项目要求：请采用 Adaboost 模型，对“classification”目录下的数据进行建模。Adaboost 模型，需自己确定并编程实现。不能采用编程语言或第三方现有编程模型。基本分类器可采用采用编程语言或第三方现有编程模型。

数据集有三个数据文件，均为 csv 格式，其中训练集有 10000 个样本，验证集和测试集均有 1000 个样本。训练集和验证集包含了标签信息，用于同学们训练和验证模型；测试数据集用于评价大家模型的学习性能。请同学提交训练结果时，将测试集的预测结果写入“adaboost_classification_test.csv”文件的 y 列中，分类结果为 0 或 1。

编程语言：自选，建议采用 Matlab 或 Python。

项目评分：分类数据集采用正确率作为评价指标，以及源文件的可读性作为评价指标。

提交内容：程序源文件，“adaboost_classification_test.csv”文件。