第1章 机器学习概念

Exercise 3.1

关于线性回归的描述，以下正确的有?

A、基本假设包括随机干扰项是均值为0，方差为1的标准正态分布
B、基本假设包括随机干扰项是均值为0的同方差正态分布
C、线性回归常用来预测离散的独立变量
D、在违背基本假设时，模型不再可以估计

Exercise 3.2

After getting $\hat{w}$, we can calculate the predictions $\hat{y}_n = w^T \mathbf{x_n}$. If all $\hat{y}_n$ are collected in a vector $\hat{Y}$ similar to how we form Y, what is the matrix formula of $\hat{y}$?

A、Y
B、XX^TY
C、$XX^\dagger Y$
D、$XX^\dagger XX^TY$

Exercise 3.3

Consider using linear regression hypothesis $h(\mathbf{x}) = w^T\mathbf{x}$ to predict the credit limit of customers $\mathbf{x}$. Which feature below shall have a positive weight in a good hypothesis for the task?

A、birth month
B、monthly income
C、current debt
D、number of credit cards owned

Exercise 3.4

The weight update rule in formula $w(t+1) = w(t) + y(t)x(t)$ has the nice interpretation that it moves in the direction of classifying $x(t)$ correctly.

(a) Show that $y(t)w^T(t)x(t) < 0$. [Hint: $x(t)$ is misclassifed by $w(t)$.]
(b) Show that $y(t)w^T(t+1)x(t) > y(t)w^T(t)x(t)$.
(c) As far as classifying $x(t)$ is concerned, argue that the move from $w(t)$ to $w(t+1)$ is a move 'in the right direction ' .

Exercise 3.5

现有一个线性回归预测函数为：

$$f(x) = w_1 \cdot x_1 + w_2 \cdot x_2 + b$$

如果我们的目标是最小化$f(x)$和y(真实值)的均方误差:

$$\arg \min_{w_1, w_2, b} \frac{1}{N} \sum_{i=1}^{N} (f(x_i) - y_i)^2$$

试求出此时的w_1和b。

Exercise 3.6

已知一个训练数据集，其正实例点x1 = (2,4)，x2 =(3,3) 负实例点是 x3 = (0,1)，试用感知机学习算法，求感知机模型$f(x) = sign(w \cdot x + b)$ (注每次的学习率为0.5)，其中损失函数为均方差。

注：按照感知机算法给出每次过程

实践题：

Exercise 3.1 房价预测：

已有某市真实房价销售数据，该数据集包含了8个特征以及一个目标数据价格。

8个特征分别为：

longitude 经度

latitude 维度

housing_median_age 街区平均房龄

total_rooms 街区总房数

total_bedrooms 街区总卧室

population 街区人口

households 街区住户

median_income 收入中位数

ocean_proximity 离海距离

预测目标：

median_house_value 房价中位数

编程实现线性回归算法，通过学习此数据集来预测房价中位数。最后提交相应的代码和在测试集上的准确率。

数据集地址：

https://query.data.world/s/yffqqcx3rsjlzspztxr6zt5iqd45kn

注： 拿到数据集后, 先按8:2划分数据集分别形成训练集和测试集。