

第四次作业

Exercise 4.1

下列关于线性回归分析中的残差 (Residuals) 说法正确的是？

- A. 残差均值总是为零
- B. 残差均值总是小于零
- C. 残差均值总是大于零
- D. 以上说法都不对

Exercise 4.2

假如我们使用 Lasso 回归(L1正则化)来拟合数据集, 该数据集输入特征有 100 个(x_1, x_2, \dots, x_{100})。现在, 我们把其中一个特征值扩大 10 倍 (例如是特征 x_1), 然后用相同的正则化参数对 Lasso 回归进行修正。

那么, 下列说法正确的是？

- A. 特征 x_1 很可能被排除在模型之外
- B. 特征 x_1 很可能还包含在模型之中
- C. 无法确定特征 x_1 是否被舍弃
- D. 以上说法都不对

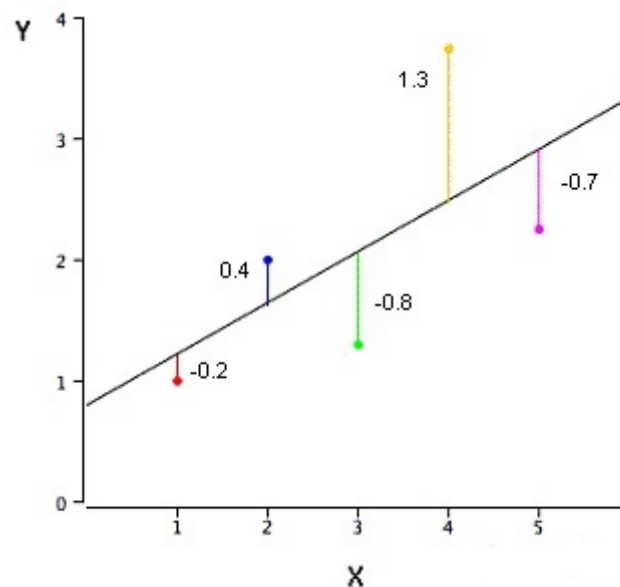
Exercise 4.3

构建一个最简单的线性回归模型需要几个系数？

- A. 1 个
- B. 2 个
- C. 3 个
- D. 4 个

Exercise 4.4

下面这张图是一个简单的线性回归模型,图中标注了每个样本点预测值与真实值的残差。计算 SSE 为多少？



- A. 3.02
- B. 0.75
- C. 1.01
- D. 0.604

Exercise 4.5

下列关于极大似然估计 (Maximum Likelihood Estimate, MLE) , 说法正确的是 (多选) ?

- A. MLE 可能并不存在
- B. MLE 总是存在
- C. 如果 MLE 存在, 那么它的解可能不是唯一的
- D. 如果 MLE 存在, 那么它的解一定是唯一的**

Exercise 4.6

给出噪声分布符合0均值拉普拉斯分布, 模型先验服从均值为0的高斯分布所对应的损失函数的形式?

Exercise 4.7

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 misclassified 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 4.8

已知一个训练数据集, 其正实例点 $x_1 = (2, 4)$, $x_2 = (3, 3)$; 负实例点是 $x_3 = (0, 1)$, 试用感知机学习算法, 求感知机模型 $f(x) = \text{sign}(w \cdot x + b)$ (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

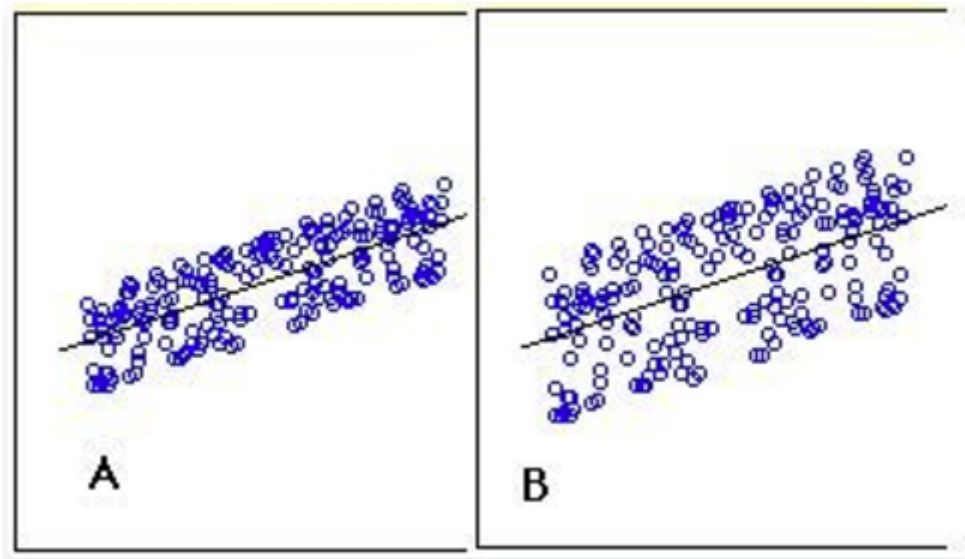
Exercise 4.9

以下关于sigmoid函数的优点说法错误的是?

- A. 可以压缩数据值到(0,1)之间, 便于后续处理
- B. 可以用于处理二分类问题
- C. 函数处处连续, 便于求导
- D. 在深层次神经网络反馈传输中, 不易出现梯度消失

Exercise 4.10

下面两张图展示两个拟合回归线 (A 和 B) , 原始数据是随机产生的。现在, 我想要计算 A 和 B 各自的残差之和。注意: 两种图中的坐标尺度一样。



关于 A 和 B 各自的残差之和，下列说法正确的是？

- A. A 比 B 高
- B. A 比 B 小
- C. A 与 B 相同
- D. 以上说法都不对

Exercise 4.11

一监狱人脸识别准入系统用来识别待进入人员的身份，此系统一共包括识别4种不同的人员：狱警，小偷，送餐员，其他。下面哪种学习方法最适合此种应用需求：

- A. 回归问题
- B. 二分类问题
- C. 多分类问题
- D. 聚类问题

Exercise 4.12

以下关于分类问题的说法错误的是？

- A. 回归问题在一定条件下可被转化为多分类问题
- B. 分类问题输入属性必须是离散的
- C. 分类属于监督学习
- D. 多分类问题可以被拆分为多个二分类问题

Exercise 4.13

以下关于逻辑回归与线性回归问题的描述错误的是

- A. 逻辑回归一般要求变量服从正态分布，线性回归一般不要求
- B. 逻辑回归用于处理分类问题，线性回归用于处理回归问题
- C. 线性回归计算方法一般是最小二乘法，逻辑回归的参数计算方法是似然估计法。
- D. 线性回归要求输入输出值呈线性关系，逻辑回归不要求

Exercise 4.14

假设有三类数据，用OVR方法需要分类几次才能完成？

- A. 2
- B. 3
- C. 1
- D. 4

Exercise 4.15

逻辑回归的损失函数是哪个？

- A. MAE B. RMSE C. MSE D. 交叉熵(Cross-Entropy)损失函数

Exercise 4.16

你正在训练一个分类逻辑回归模型。以下哪项陈述是正确的？

- A. 将正则化引入到模型中，总是能在训练集上获得相同或更好的性能
B. 向模型中添加新特征总是会在训练集上获得相同或更好的性能
C. 将正则化引入到模型中，对于训练集中没有的样本，总是可以获得相同或更好的性能
D. 在模型中添加许多新特性有助于防止训练集过度拟合

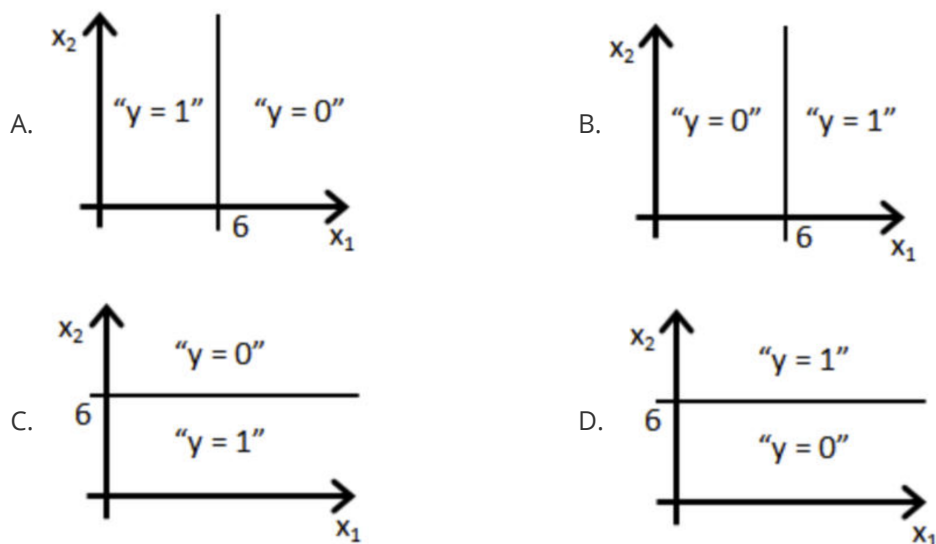
Exercise 4.17

假设您进行了两次逻辑回归，一次是 $\lambda=0$ ，一次是 $\lambda=1$ （ λ 是正则化参数）。其中一次，得到参数 $w=[81.47, 12.69]$ ，另一次，得 $w=[13.0, 10.91]$ 。但是，您忘记了哪个 λ 值对应于哪个 w 值。你认为哪个对应于 $\lambda=1$ ？

- A. $w=[13.0, 10.91]$
B. $w=[81.47, 12.69]$

Exercise 4.18

假设训练一个逻辑回归分类器 $h_w(\mathbf{x}) = \theta(w_0 + w_1x_1 + w_2x_2)$ 。假设 $w_0 = 6, w_1 = -1, w_2 = 0$ ，下列哪个图表示分类器找到的决策边界？



Exercise 4.19

[不定项选择题] 假设您有以下训练集，并拟合logistic回归分类器

$$h_w(\mathbf{x}) = g(w_0 + w_1x_1 + w_2x_2)$$

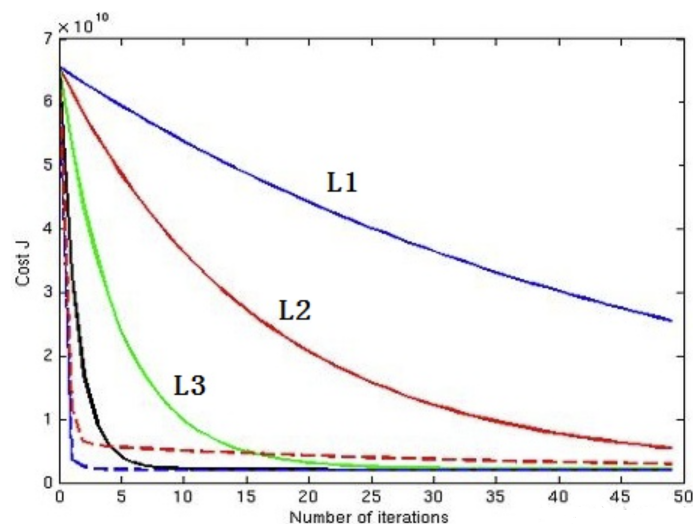
x_1	x_2	y
1	0.5	0
1	1.5	0
2	1	1
3	1	0

以下哪项是正确的？选出所有正确项

- A. 添加多项式特征（例如，使用 $h_{\mathbf{w}}(\mathbf{x}) = g(w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_1x_2 + w_5x_2^2)$ ）可以增加我们拟合训练数据的程度
- B. 在 \mathbf{w} 的最佳值处， $J(\mathbf{w}) \geq 0$
- C. 添加多项式特征（例如，使用 $h_{\mathbf{w}}(\mathbf{x}) = g(w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_1x_2 + w_5x_2^2)$ ）将增加 $J(\mathbf{w})$ ，因为我们现在正在对更多项进行求和
- D. 如果我们训练梯度下降迭代足够多次，对于训练集中的一些例子 \mathbf{x}_i ，可能得到 $h_{\mathbf{w}}(\mathbf{x}_i) > 1$

Exercise 4.20

如图显示了逻辑回归中3种不同学习速率值的代价函数和迭代次数之间的关系，L_1、L_2、L_3为对应的学习速率，下面哪一个选项是正确的？



- A、 $L1 > L2 > L3$ B、 $L1 = L2 = L3$
- C、 $L1 < L2 < L3$ D、都不是

Exercise 4.21

为什么要使用：

- a. 岭回归而不是简单的线性回归（即没有任何正则化）？
- b. Lasso而不是岭回归？
- c. 弹性网络而不是Lasso？

Exercise 4.22

训练逻辑回归模型时，梯度下降会卡在局部最小值中吗？