

人工智能基础作业 1

林乐天 — 2300012154@stu.pku.edu.cn

2024 年 3 月 25 日

1 问答 1

1.1 问题

请简述什么是贝叶斯定理，什么是最大似然估计 (MLE)，什么是最大后验估计 (MAP)。

1.2 答案

贝叶斯定理 贝叶斯定理是一个描述对事件发生可能性的信心的公式，基于对已有可能与事件相关的先验知识来描述时间发生的可能性。其公式为： $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ ，称 $P(B|A)$ 为先验概率，称 $P(A|B)$ 为后验概率。

最大似然估计 最大似然估计是一种优化方法，是在假定数据随机无偏的情况下，通过使得 $P(B|A)$ 最大的方法，使得模型最大限度的拟合现实情况，对数据集的要求很高。

最大后验估计 最大后验估计也是一种优化方法，本质上是不信任数据集能真实反映现实情况，通过假设一个先验分布，最大化 $P(A|B)$ 即最大化 $P(B|A) * P(A)$ 来优化模型的方法，对先验分布比较敏感。

2 问答 2

2.1 问题

设 $X \sim N(\mu, \sigma^2)$, μ, σ^2 为未知参数, $x_1 \dots x_n$ 是来自 X 的样本值, 求 μ, σ^2 的最大似然估计量。

2.2 答案

$$\mu = \frac{1}{n} \sum_i x_i = \bar{x}$$
$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

3 问答 3

3.1 问题

请简述分类问题与回归问题的主要区别。

3.2 答案

回归问题输出的是一个连续值, 是对一个连续函数的预测和拟合, 其标签是连续的; 而分类问题输出的是一个离散值, 是对一个离散值的预测, 其标签是离散的。

4 问答 4

4.1 问题

请简述有监督学习与无监督学习的主要区别。

4.2 答案

有监督学习除了给出数据, 还给出了 label; 无监督学习只给出数据, 不给出 label。

5 问答 5

5.1 问题

给定数据 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 用一个线性模型估计最接近真实 y (ground truth) 的连续标量 Y , $f(x_i) = Wx_i + b$, such that $f(x) \approx y$. 求最优 (w^*, b^*) 使得 $f(x)$ 与 y 之间的均方误差最小并解释 (w^*, b^*) 何时 closed form 解, 何时没有 closed form 解。

5.2 答案

先写出均方误差:

$$L(W, b) = \frac{1}{n} \sum Wx_i + b - y_i$$

$$L(\beta) = \frac{1}{n} (A\beta - Y)^T (A\beta - Y)$$

其中, A 包含了数据的信息, 且加了一行 1; β 是前文 W 和 b 的组合。

对 β 求偏导并求零点, 化简得: $(A^T A)\beta = A^T Y$

如果 $(A^T A)$ 可逆, 则其有 closed form 的解, 否则没有。

6 问答 6

6.1 问题

Ridge regression 问题的解具有什么特点, 为什么? Lasso 问题的解具有什么特点? 为什么?

6.2 答案

Ridge regression 这种方法的解相对来说其绝对值较小。由于其为 L_2 距离, 对绝对值大小比较敏感, 且容易在非坐标轴的位置和解域相交, 所以绝对值小。

Lasso regression 这种方法的解相对维度较小, 会出现大量的 0。由于其为 L_1 距离, 解域很容易在坐标轴上相交, 于是如此。

7 问答 7

7.1 问题

请从 model function、loss function、optimization solution 三个方面比较 Linear regression 与 Logistic regression 的异同。

7.2 答案

model function 对于 Linear 来说, model function 只是一个单纯的对 X 的 linear combination, 在加上一个 bias, 是一个纯粹的线性形式 $y = Wx + b$

而对于 Logistic regression 来说, 为了让结果可以解释, 要在线性组合外面套一个 sigmoid 函数, 就使得整个函数并非完全的线性, 输出值也被控制在了 $[0,1]$ 上。

loss function 对于 Linear 来说, 更常用的是基本的 Square Loss, 由于取值范围相对较大, 使用均方误差收敛速度尚可, 故如此。

而对于 Logistic 来说, 若使用 Square Loss, 会在结果严重偏离正确结果的时候不能收敛, 其梯度会变成 0, 于是常用 Cross Entropy Loss, 从而保证其在梯度下降的时候可以正常运作。

optimization solution 二者在这个方面具有一致性。两个函数都是用梯度下降的方法来进行优化的, 所以二者在这个方面是相同的。

8 问答 8

8.1 问题

K-近邻分类器的超参数是什么? 怎么选择 K-近邻分类器的超参数?

8.2 答案

超参数 其超参数有两个, 一个是 K 的值, 另一个是距离量度的选择。

选择 通过在测试集上进行调试找到合适的超参数, 在数据集不够大的时候可以进行交叉验证。