蒋豆芽

机器学习——线性回归与逻辑回归 已购

来自 【机器学习面试题汇总与解析（蒋豆芽面试题总结）】 | 61 浏览 | 0 回复 | 2021-05-22

蒋豆芽

+关注

机器学习面试题汇总与解析——线性回归与逻辑回归

1. **逻辑回归 LR 详细推导**☆☆☆☆☆
2. **回归和分类的区别**☆☆☆☆☆
3. **逻辑回归特征是否归一化**☆☆☆☆☆
4. **什么样的模型需要特征归一化**☆☆☆☆☆
5. **如何提升LR的模型性能?** ☆☆☆☆☆
6. **逻辑回归为啥要做特征离散化**☆☆☆☆☆
7. **LR的详细过程，如何优化**☆☆☆☆☆
8. **知道什么损失函数，lr公式推导**☆☆☆☆☆
9. **最小二乘法在什么条件下与极大似然估计等价?** ☆☆☆☆☆
10. **逻辑回归为什么不用平方损失函数?** ☆☆☆☆☆

==========

- 本专栏适合于**Python已经入门**的学生或人士，有一定的编程基础。
- 本专栏适合于**算法工程师、机器学习、图像处理求职**的学生或人士。
- 本专栏针对面试题答案进行了**优化，尽量做到好记、言简意赅。这才是一份面试题总结的正确打开方式。这样才方便背诵**
- 如专栏内容有错漏，欢迎在评论区指出或私聊我更改，一起学习，共同进步。
- **相信大家都有着高尚的灵魂，请尊重我的知识产权，未经允许严禁各类机构和个人转载、传阅本专栏的内容。**

==========

关于**机器学习算法**书籍，我强烈推荐一本**《百面机器学习算法工程师带你面试》**，这个就很类似面经，还有讲解，写得比较好。私聊我进群。

关于**深度学习算法**书籍，我强烈推荐一本**《解析神经网络——深度学习实践手册》**，简称CNN book，通俗易懂。私聊我进群。

蒋豆芽

B站机器学习视频：https://space.bilibili.com/1078114/channel/detail?cid=139904

线性回归：https://zhuanlan.zhihu.com/p/90998021

逻辑回归：https://zhuanlan.zhihu.com/p/51279024

逻辑回归实现：https://zhuanlan.zhihu.com/p/55438631

线性回归与逻辑回归公式推导：https://blog.csdn.net/qq_37537170/article/details/107388449

读者可以把参考文章看看

个人理解

1. **线性回归是回归（预测），逻辑回归是分类。**
2. **线性回归，输出套上sigmoid函数就成了逻辑回归**

两者优缺点

优点

1. 模型简单，原理简单易理解
2. 计算代价不高，易于理解和实现。

缺点：

1. 易过拟合
2. 特征很多的时候，效果不好
3. 处理线性问题效果比较好，而对于更复杂的问题可能束手无策

1. **逻辑回归 LR 详细推导**☆☆☆☆☆

   参考回答

   分为两个部分，一是推导逻辑回归的公式，二是推导损失函数。直接参考文章内容。

   答案解析

   无。

类似的问题还有：

2. **回归和分类的区别**☆☆☆☆☆

   参考回答

   1. 两者的的预测目标变量类型不同，回归问题是**连续变量**，分类问题**离散变量**。

4. 评价指标不用：回归的评价指标通常是MSE；分类评价指标通常是Accuracy、Precision、Recall

答案解析

无。

3. **逻辑回归特征是否归一化** ☆ ☆ ☆ ☆ ☆

参考回答

逻辑回归本身不受**量纲**影响，但是其使用**梯度下降法**求解参数受**量纲**影响大，如果不进行特征归一化，可能由于变量不同量纲导致参数迭代求解缓慢，影响算法速率。

答案解析

一般算法如果本身受量纲影响较大，或者相关优化函数受量纲影响大，则需要进行**特征归一化。**逻辑回归本身不受量纲影响，但是其使用梯度下降法求解参数受量纲影响大，如果不进行特征归一化，可能由于变量不同量纲导致参数迭代求解缓慢，影响算法速率。

对于决策树这类的算法，不受**量纲**影响，不需要进行归一化处理。

4. **什么样的模型需要特征归一化** ☆ ☆ ☆ ☆ ☆

参考回答

一般算法如果本身受量纲影响较大，或者相关优化函数受量纲影响大，则需要进行特征**归一化。**

答案解析

无。

5. **如何提升LR的模型性能?** ☆ ☆ ☆ ☆ ☆

参考回答

1. 想办法获得或构造更多的数据，无论评估模型还是训练，都会更加可靠。
2. 根据业务知识，挖掘更多有价值的Feature，即特征工程。
3. 加入正则化项，L1/L2。交叉验证确定最优的参数。这会加快模型开发速度，会自动化筛选变量。

答案解析

蒋豆芽

6. **逻辑回归为啥要做特征离散化** ☆ ☆ ☆ ☆ ☆

   参考回答

   1. **非线性**：逻辑回归属于广义线性模型，表达能力受限；单变量离散化为N个后，每个变量有单独的权重，相当于为模型引入了非线性，能够提升模型表达能力，加大拟合； 离散特征的增加和减少都很容易，易于模型的快速迭代；
   2. **速度快**：稀疏向量内积乘法运算速度快，计算结果方便存储，容易扩展；
   3. **鲁棒性**：离散化后的特征对异常数据有很强的鲁棒性：比如一个特征是“年龄>30是1，否则0”。如果特征没有离散化，一个异常数据“年龄300岁”会给模型造成很大的干扰；
   4. **方便交叉与特征组合**：离散化后可以进行特征交叉，由M+N个变量变为M*N个变量，进一步引入非线性，提升表达能力；
   5. **简化模型**：特征离散化以后，起到了简化了逻辑回归模型的作用，降低了模型过拟合的风险。

   答案解析

   无。

7. **LR的详细过程，如何优化** ☆ ☆ ☆ ☆ ☆

   参考回答

   答案参考文章内容。

   答案解析

   无

8. **知道什么损失函数，lr公式推导** ☆ ☆ ☆ ☆ ☆

   参考回答

   答案参考文章内容。

   答案解析

   无

9. **最小二乘法在什么条件下与极大似然估计等价？** ☆ ☆ ☆ ☆ ☆

   参考回答

   当模型估计值和真实值间的残差项服从**均值是0的高斯分布**时，就有最小二乘估计和最大似然估计等价。

加上高斯噪声 $\varepsilon \sim N(0, \sigma^2)$ 得到的，即：

$$Y = f_\theta(X) + \varepsilon$$

那么对模型参数 θ 的最大似然估计和最小二乘估计是等价的。

----------简单的推导----------

我们知道，模型的似然函数是

$$L(\theta) = \log P(Y|X, \theta) = \sum_i \log P(y_i|x_i, \theta)$$

同时，有 $y_i \sim N(f_\theta(x_i), \sigma^2)$，那么可以得到

$$L(\theta) = -\frac{1}{2\sigma^2}\sum_i (y_i - f_\theta(x_i))^2 - N\log\sigma - \frac{N}{2}\log 2\pi$$

因此，去掉后面两项不包含 θ 的常数项，模型参数 θ 的最大似然估计 $\max_\theta L(\theta)$，就等价于最小二乘估计 $\min_\theta \sum_i (y_i - f_\theta(x_i))^2$。

10. **逻辑回归为什么不用平方损失函数?** ☆☆☆☆☆

参考回答

因为逻辑回归的平方损失函数是非凸函数，梯度下降时很难得到全局极值点。

答案解析

无

资源分享 python 机器学习 算法工程师 春秋招 面试题 软件开发 面经 举报

收藏 赞

相关专栏

蒋豆芽

0条评论 默认排序

没有回复

请留下你的观点吧~

发布