

自然语言处理

杨海钦
2025-2026-1 学期
基于Yulia, Graham, Diyi, Jixing等课件更新

面试问题

- 假设你需要为公司训练一个“情感分类器”（输入一段文本，输出“积极”、“消极”或“中性”），请回答以下问题：
 - 如果你只可以选择朴素贝叶斯、逻辑回归、结构化感知器三种分类器，你会如何选择？
 - 在实现上述模型的时候，你如何解决下述问题？
 - 朴素贝叶斯: 如何处理浮点下溢？如何解决训练数据的词没有
 - 多项式逻辑回归: 如何避免数据计算溢出
 - 如何获得优化性能？

推荐设置

- Conda + VS Code + Copilot

问题

- 如何从Jupyter Note自动输出.py文件？

实现多项式逻辑回归

- 补全代码：
 - 如何提特征？
 - 如何建立字典库(Bag of Words)？
 - 如何从文本转数值？
 - 如何实现softmax(避免数值溢出)
 - 如何实现交叉熵损失
 - 如何训练模型？
 - 向前传播
 - 计算梯度
 - 更新梯度
- https://github.com/hqyang/nlp-codes/blob/main/01-simpleclassifier/train_logistic_regression.ipynb
- https://www.educoder.net/classrooms/99677/common_homework/