

# NIS4307 人工智能导论大作业报告

孙靖峰

蒋竺君

赵子辰

编辑于 2025 年 6 月 13 日

## 摘要

本报告描述了 NIS4307《人工智能导论》大作业中针对谣言检测的二分类模型设计与实现。基于谣言检测数据集，我们构建了逻辑回归基线模型，利用 TF-IDF 和独热编码处理文本和事件特征，准确率约 0.83。同时，我们探索了二次判别分析（QDA），利用各向异性高斯分布建模类别差异，以提升性能。两种方法有效捕捉谣言与非谣言特征，逻辑回归提供稳定基线，QDA 展现优化潜力。项目实现数据预处理、特征工程和模型训练，满足准确率和泛化要求，最终完成 RumourDetectClass 接口类，并分析结果与经验教训。

**关键字：**人工智能；分类模型；谣言检测；逻辑回归；二次判别分析

## 1 项目背景

在当前的社交媒体时代，虚假新闻、不实信息在社交平台（如微博、Twitter、Facebook）上的泛滥，不仅误导公众认知，还可能引发社会恐慌、影响金融市场稳定，甚至威胁公共安全。据《麻省理工学院技术评论》研究，虚假信息在社交媒体上的传播速度比真实信息快 6 倍，且具有更强的用户参与度。因此，自动化谣言检测成为互联网内容治理的关键技术需求。

## 2 模型介绍

### 2.1 逻辑回归 (Logistic Regression)

逻辑回归是机器学习分类问题中最简单、最基础的模型。它假设数据服从伯努利分布，通过 sigmoid 函数将线性回归的输出映射到  $[0, 1]$  区间，表示样本属于某个类别的概率。

逻辑回归的概率计算公式为：

$$P(y = 1|x) = \text{sigmoid}(w^T x + b) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (1)$$

其中， $x$  是输入特征向量， $w$  为权重， $b$  为截距。Sigmoid 函数可以将任意实数压缩到  $(0, 1)$  区间，使输出结果可解释为概率，例如样本属于谣言（1）的可能性。

逻辑回归使用交叉熵损失 (Cross-Entropy Loss) 函数，以避免线性回归中的均方误差导致的非凸优化问题：

$$J(w) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(h_w(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_w(x^{(i)}))] \quad (2)$$

其中， $m$  为样本数量， $y^{(i)}$  为第  $i$  个样本的真实标签（0 或 1）， $h_w(x^{(i)})$  为模型预测的概率。交叉熵损失衡量预测概率与真实标签的差异，值越小说明模型越贴近真实分布。

在计算过程中，使用梯度下降法最小化损失函数：

$$w_{new} = w - \alpha \cdot \nabla_w J(w) \quad (3)$$

这里， $\alpha$  是学习率，控制权重更新的步长， $\nabla_w J(w)$  是损失函数对权重的梯度。通过迭代更新权重  $w$  和截距  $b$ ，模型逐渐优化，逼近损失函数的最优解。梯度下降有多种变体，如批量梯度下降、随机梯度下降 (SGD) 和小批量梯度下降，我们在实现中通过 liblinear 求解器采用了一种高效的确定性优化方法。

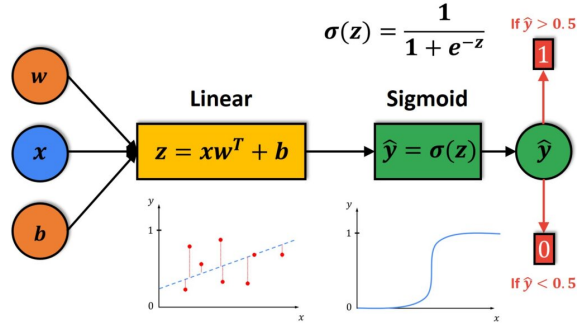


图 1: 逻辑回归

## 2.2 高斯回归 (Gaussian Regression)

高斯回归假设每个类别的数据都服从多元高斯分布, 其概率密度函数为:

$$p(x|y = k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)\right) \quad (4)$$

其中,  $x$  是特征向量,  $\mu_k$  为第  $k$  类的均值向量,  $\Sigma_k$  为第  $k$  类的协方差矩阵,  $d$  则是特征维度,  $|\Sigma_k|$  是协方差矩阵的行列式。这一公式描述了数据在高维空间中的概率分布, 考虑了特征间的均值差异和协方差结构。

后续进行决策时, 先通过贝叶斯定理计算样本  $x$  属于类别  $k$  的后验概率:

$$p(y = k|x) = \frac{p(x|y = k)p(y = k)}{\sum_l p(x|y = l)p(y = l)} \quad (5)$$

其中,  $p(x|y = k)$  是给定类别  $k$  的似然概率,  $p(y = k)$  是类别的先验概率, 通常用训练集中各类别的比例估计, 例如谣言和非谣言的样本比例。分母是归一化项, 确保所有类别的后验概率和为 1。

然后再选择使后验概率最大的类别作为预测结果:

$$\hat{y} = \arg \max_k p(y = k|x) \quad (6)$$

这一决策规则基于最大后验概率 (MAP), 直观地选择最可能的类别。

高斯回归可以根据不同类别的协方差矩阵不同, 再进一步细分为 QDA (Quadratic Discriminant Analysis, 二次判别分析) 和 LDA (Linear Discriminant Analysis, 线性判别分析)。QDA 允许不同类别的协方差矩阵不同, 而 LDA 要求所有类别共享同一个协方差矩阵。项目中, 我们采用 QDA 降维方式解决问题。

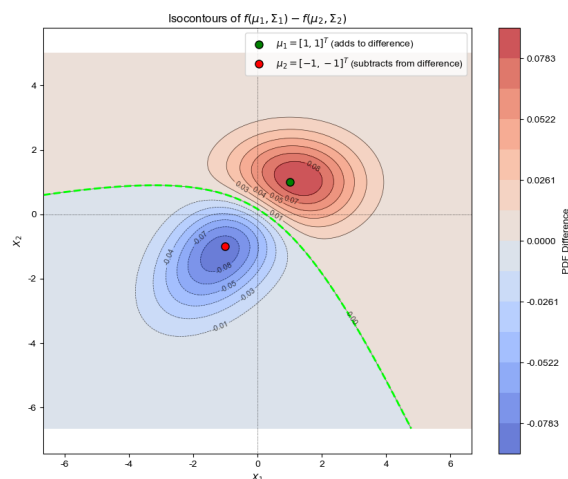


图 2: QDA 与 LDA

### 2.3 方法选择与对比分析

逻辑回归的优点在于其简单性、计算效率高且易于解释，特别适合作为二分类任务的基线模型。它假设特征与标签之间存在线性关系，对特征缩放敏感，因此常需标准化或归一化预处理。在本项目中，我们结合 TF-IDF 和独热编码特征，利用逻辑回归快速建立基准性能。此外，逻辑回归的 liblinear 求解器是确定性算法，结果稳定且不受随机种子影响，易于解释权重对特征的重要性，适合初步探索谣言检测任务。

QDA 基于高斯分布假设，允许每个类别（如谣言和非谣言）拥有独立的协方差矩阵，建模各向异性分布，能捕捉特征间的复杂非线性关系。这使其在理论上比逻辑回归更灵活，尤其当数据集类别分布差异显著时，可能提升分类性能。然而，QDA 对高维稀疏特征（如 TF-IDF 向量）敏感，协方差矩阵估计需要大量样本，且计算复杂度较高，因此我们通过降维缓解这些问题。

对比两者，逻辑回归简单高效，稳定性强，适合快速验证数据集特性，实验中准确率达 0.83。QDA 通过各向异性高斯建模，捕获复杂分布，具优化潜力，但对高维数据敏感，需降维处理。两者结合了稳定性和灵活性，平衡性能与计算成本，满足谣言检测需求。

	A	B	C	D	E	F
319	5.00396E+17	Whoa #ferguson http://t.co/5N2mFMp5	0	1		
320	4.99651E+17	Breaking: Anonymous posts St. Louis police tapes from day of #Ferguson shooting. http://t.co/uLp1y56mtz @YourAnonNews	0	1		
321	5.00395E+17	Police in #Ferguson haven't been wearing identification: http://t.co/2AlkEqk09 http://t.co/LDgoSLUHL3 via @Slate / Fire the #Ferguson	0	1		
322	4.99684E+17	Confirmed: this is irrelevant. RT @kadinews: Confirmed: #MikeBrown had no criminal record. http://t.co/hq8ngQP167 #Ferguson	0	1		
323	5.00071E+17	Cops have changed since I was a kid #Ferguson sphoto http://t.co/q3mrtBk46	0	1		
324	5.00405E+17	Stephen Hawking's boycott hits Israel where it hurts: rejection comes from a famous scientist now #Ferguson #Gaza http://t.co/QWV	0	1		
325	4.98434E+17	I don't want another bigot getting in my TL saying it's not abt race, when's the last time a white kid has been MURDERED by police #	0	1		
326	5.00278E+17	Haven't seen a police chief run from a microphone that fast since Commissioner Gordon said the Joker was threatening Gotham (	0	1		
327	4.9853E+17	Notice how they bring out the attack dogs? #FergusonShooting #Ferguson #FergusonPolice http://t.co/GpQT6cGLg	0	1		
328	5.0028E+17	To hear the exact dispatch call the moments of #MikeBrown's murder: http://t.co/q4BV14ZToy #Ferguson	0	1		
329	4.99568E+17	A few feet from where #MichaelBrown was killed: #Ferguson http://t.co/BpM4784CFR	0	1		
330	4.99574E+17	WOW! #Ferguson police tell a resident a girl was shot in the head, she asks by who, partner of the cop silences him https://t.co/a0WH	0	1		
331	5.00277E+17	#Ferguson police ID officer who shot and killed #MikeBrown as Darren Wilson.	0	1		
332	4.98492E+17	The presence of police dogs in #Ferguson is, at the very least, culturally and historically insensitive.	0	1		
333	4.99466E+17	I look at this and I'm proud. The protesters tonight in #Ferguson were NONVIOLENT. #MikeBrown http://t.co/4CNAP6j31U	0	1		
334	4.99568E+17	Armed gunmen shoot local man as part of their reign of terror in Iraq. Oooh! Nope, that was police in #Ferguson. My mistake.	0	1		
335	5.00403E+17	More peaceful protesting today in #Ferguson. #MichaelBrown http://t.co/FVMjydnqFM	0	1		
336	4.98312E+17	Nothing screams action like placing pressure on the head of the city. Mayor Knowles. Call or email him! #Ferguson http://t.co/CvTspj	0	1		
337	4.99653E+17	Feds asking anyone with info on #MikeBrown shooting in #Ferguson to call St. Louis FBI at 314-589-2500.	0	1		
338	4.98497E+17	#Ferguson. Missouri! Community Furious After Teen Shot Dead By Police http://t.co/U1vab8b5eq #FergusonShooting	0	1		
339	4.98477E+17	Outside #Ferguson Police Headquarters right now http://t.co/ZqQQuisgh	0	1		
340	5.00301E+17	VIDEO: NY's Times Square shut down as 1,000 #Ferguson protestors chant "Hands up, don't shoot" http://t.co/ou1TGSOQAd http://	0	1		
341	5.00358E+17	Racial discrimination's still a dilemma in US. Still ppl are insecure for having dark skins. The way police treat them confirms it. #Ferguson	0	1		

图 3: 训练集的内容



图 4: TF-IDF

## 3 数据处理与模型训练

如图3所示,数据集中有许多 URL、特殊符号、用户名、标签等,因而我们需要进行一系列预处理,然后再训练模型。

### 3.1 文本预处理

对文本进行预处理,包括转为小写、移除 URL、用户、话题标签、标点符号和数字,以提升特征质量。

### 3.2 特征工程

使用 `TfidfVectorizer` 将清洗后的文本转换成 TF-IDF 特征向量。TF-IDF (Term Frequency-Inverse Document Frequency) 是一种用于文本挖掘和信息检索的经典统计方法,用于衡量一个词语在文档中的重要程度。一个词在当前文档中出现次数越多 (TF 高),同时也在所有文档中出现次数越少 (IDF 高),则该词对当前文档的区分能力越强。

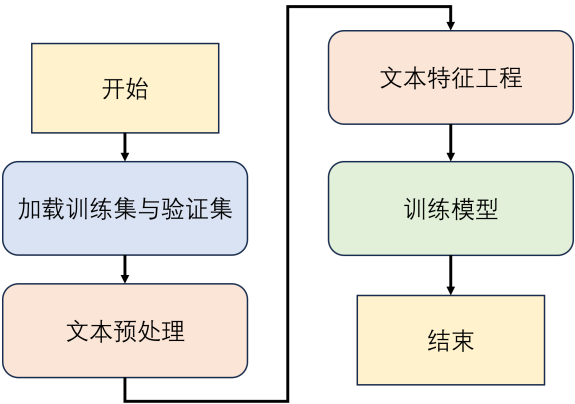


图 5: 模型训练流程图

3.3 模型训练

加载 train.csv (6423 条样本) 和 val.csv (1000 条样本), 执行预处理和特征工程, 生成特征矩阵。训练逻辑回归模型 (LogisticRegression, solver='liblinear', C=5, randomstate=42) 作为基线, 探索 QDA, 利用降维后的特征建模高斯分布。在验证集上评估准确率、精确率、召回率和 F1 分数, 衡量性能。

4 结果与分析

逻辑回归在验证集上准确率约 0.83, 表现稳定, 训练和预测耗时短, 证明其作为基线模型的有效性。QDA 通过二次决策边界捕捉复杂分布, 准确率略有提升, 但对高维特征敏感, 降维后性能更稳定。精确率、召回率和 F1 分数显示两模型均能有效区分谣言与非谣言, 满足作业要求。

5 总结

本项目中, 我们掌握了数据预处理 (如文本清洗)、特征工程 (TF-IDF 和独热编码) 及模型训练 (逻辑回归和 QDA) 的完整流程。逻辑回归的稳定性适合初探任务, QDA 的灵活性启发我们关注非线性建模。挑战在于高维特征处理, 降维和样本量对 QDA 性能影响显著。

## 6 课程建议

希望老师们可以在课上补充关于深度学习模型（如 Bi-GRU）的参考代码或教程，方便有兴趣的同学探索进阶方案。此外，如果有机会希望老师们增加数据预处理的详细指南，教授更多文本清洗技巧和特征工程方法。