

Multimodal Model Homework 5

2023141460251 孙浩翔

问题 1：单模态表示有哪些？

1.1 手工设计特征

依赖领域先验，如图像的 SIFT/HOG/LBP。

$$\theta(x, y) = \arctan \frac{I_y}{I_x} \quad (1)$$

$$h_k = \sum_{(x,y) \in \mathcal{N}} w(x, y) \mathbf{1}(\theta(x, y) \in \text{bin}_k) \quad (2)$$

- SIFT：关键点检测 + 局部梯度方向直方图；
- HOG：固定网格 + 梯度方向统计，常用于行人检测；
- LBP：局部二值模式，计算邻域灰度差异的二进制编码。

1.2 统计特征

文本和信号常用 BoW、TF-IDF、话题模型。

$$\text{tfidf}_{i,j} = \frac{f_{i,j}}{\sum_k f_{k,j}} \cdot \log \frac{N}{\text{df}_i} \quad (3)$$

$$\text{df}_i = |\{j : f_{i,j} > 0\}| \quad (4)$$

- Bag-of-Words：词频直方图，不考虑词序；
- TF-IDF：抑制高频通用词，强调区分性高的关键词；
- LDA 主题：假设文档由多主题混合生成，学习主题-词分布。

1.3 深度学习特征

通过端到端网络自动学习。

$$\mathbf{z}^{(l)} = \sigma(W^{(l)}\mathbf{z}^{(l-1)} + b^{(l)}) \quad (5)$$

$$\mathbf{z}^{(0)} = \mathbf{x}, \quad l = 1, \dots, L \quad (6)$$

- CNN: 局部感受野 + 权重共享;
 - RNN/Transformer: 时序和上下文建模;
 - 自监督: SimCLR/MoCo 通过对比学习获得泛化特征。
-

问题 2: 数据集一般由哪些组成? 分别的作用是什么?

监督学习数据集主要分为三部分:

2.1 训练集

用于模型参数学习, 通常占比 60%-80%, 通过最小化经验风险学习最优参数:

$$\min_{\theta} \frac{1}{|\mathcal{D}_{\text{train}}|} \sum_{(\mathbf{x}, y) \in \mathcal{D}_{\text{train}}} L(y, f_{\theta}(\mathbf{x})) \quad (7)$$

- 支持向量机等需要大量样本以稳定估计超平面;
- 深度模型依赖大规模训练集避免过拟合;

2.2 验证集

用于超参数调优和早停, 通常占比 10%-20%。

- 监控验证损失或 F_1 值变化;
- 调整学习率、正则化系数、网络结构等超参数;
- 防止过拟合, 确定最佳训练轮数 (early stop)。

2.3 测试集

用于最终模型评估, 通常占比 10%-20%。

- 在训练和调参完成后使用, 不参与任何训练或选择;
 - 报告准确率、 F_1 、AUC 等泛化性能指标;
 - 保证实验结果客观可信。
-

问题 3：简单分类器是什么？简述其过程？

3.1 k-最近邻

距离度量和多数投票方法：

$$\text{dist}(\mathbf{z}, \mathbf{z}_i) = \|\mathbf{z} - \mathbf{z}_i\|_2 \quad (8)$$

$$y = \arg \max_c \sum_{i \in \mathcal{N}_k(\mathbf{z})} \mathbf{1}(y_i = c) \quad (9)$$

特点：

- 无需训练阶段，计算复杂度 $O(Nd)$ ；
- 对异常值敏感，需要合理选择 k ；

3.2 朴素贝叶斯

条件独立假设下的生成模型：

$$P(y|\mathbf{x}) \propto P(y) \prod_j P(x_j|y) \quad (10)$$

$$\hat{y} = \arg \max_y P(y|\mathbf{x}) \quad (11)$$

特点：

- 参数估计简单，适合高维稀疏数据；
- 假设强，但在文本分类中常表现良好；

3.3 决策树

基于特征划分的树状模型：

$$\text{IG}(D, a) = H(D) - \sum_{v \in \mathcal{V}} \frac{|D_v|}{|D|} H(D_v) \quad (12)$$

$$H(D) = - \sum_c p_c \log p_c \quad (13)$$

特点：

- 可解释性强，树剪枝防止过拟合；
- 易受数据噪声影响，可结合随机森林提升稳定性；

问题 4：怎么评估方法的性能？

4.1 分类评估指标

混淆矩阵基础指标：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{N} \quad (14)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (15)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

$$F_1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (17)$$

以及曲线指标：

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d\text{FPR} \quad (18)$$

4.2 回归评估指标

$$\text{MSE} = \frac{1}{N} \sum_i (y_i - \hat{y}_i)^2 \quad (19)$$

$$\text{MAE} = \frac{1}{N} \sum_i |y_i - \hat{y}_i| \quad (20)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (21)$$

$$\text{ExplainedVar} = 1 - \frac{\text{Var}(y - \hat{y})}{\text{Var}(y)} \quad (22)$$

问题 5：请简述评分函数。

5.1 线性评分函数

$$s(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b \quad (23)$$

$$\text{margin} = y \cdot s(\mathbf{x}) \quad (24)$$

用途：

- 排序与阈值决策；
- 与损失函数结合优化边界宽度；

5.2 多类评分函数

$$s_c(\mathbf{x}) = \mathbf{w}_c^\top \mathbf{x} + b_c \quad (25)$$

$$\hat{y} = \arg \max_c s_c(\mathbf{x}) \quad (26)$$

扩展：可通过 Softmax 转换为概率分布。

5.3 结构化预测评分

$$\hat{y} = \arg \max_{y \in \mathcal{Y}} [s(\mathbf{x}, y) + \Delta(y, y_i)] \quad (27)$$

问题 6：请简述线性分类器。

6.1 模型形式

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b \quad (28)$$

$$y = \text{sign}(f(\mathbf{x})) \quad (29)$$

6.2 感知机

$$\mathbf{w} \leftarrow \mathbf{w} + \eta y_i \mathbf{x}_i \quad (30)$$

$$b \leftarrow b + \eta y_i \quad (31)$$

特点：

- 在线学习，收敛性依赖于数据线性可分；
- 不产生概率输出；

6.3 Logistic 回归

$$L = - \sum_i [y_i \ln \sigma(f_i) + (1 - y_i) \ln(1 - \sigma(f_i))] \quad (32)$$

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (33)$$

优势：

- 提供概率输出；
- 损失函数凸，可全局优化；

6.4 支持向量机

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (34)$$

$$\text{s.t. } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 \quad (35)$$

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \quad (36)$$

扩展：

- 核技巧处理非线性；
- 支持向量稀疏性提高效率；

问题 7：请简述损失函数，并分别写出常用的几种损失函数。

7.1 零一损失

$$L_{0-1} = \mathbf{1}(y \neq \hat{y}) \quad (37)$$

说明：非凸，不可导，常用于理论分析。

7.2 铰链损失

$$L_{\text{hinge}} = \max(0, 1 - yf(\mathbf{x})) \quad (38)$$

说明：用于 SVM，凸但不可导于边界点。

7.3 逻辑损失

$$L_{\log} = \ln(1 + e^{-yf(\mathbf{x})}) \quad (39)$$

说明：对误差平滑惩罚，易于优化。

7.4 交叉熵损失

$$L_{\text{CE}} = - \sum_c y_c \ln p_c \quad (40)$$

$$p_c = \frac{e^{s_c}}{\sum_k e^{s_k}} \quad (41)$$

说明：用于多分类，结合 Softmax，凸可导。

问题 8：请简述人工神经网络的过程

8.1 前向传播

$$\mathbf{h}^{(0)} = \mathbf{x} \quad (42)$$

$$\mathbf{a}^{(l)} = W^{(l)}\mathbf{h}^{(l-1)} + b^{(l)} \quad (43)$$

$$\mathbf{h}^{(l)} = \sigma(\mathbf{a}^{(l)}), \quad l = 1, \dots, L \quad (44)$$

$$\hat{\mathbf{y}} = \text{softmax}(W^{(L+1)}\mathbf{h}^{(L)} + b^{(L+1)}) \quad (45)$$

8.2 反向传播

$$\delta^{(L+1)} = \hat{\mathbf{y}} - \mathbf{y} \quad (46)$$

$$\delta^{(l)} = (W^{(l+1)\top} \delta^{(l+1)}) \circ \sigma'(\mathbf{a}^{(l)}) \quad (47)$$

$$\frac{\partial L}{\partial W^{(l)}} = \delta^{(l)} \mathbf{h}^{(l-1)\top} \quad (48)$$

$$\frac{\partial L}{\partial b^{(l)}} = \delta^{(l)} \quad (49)$$

8.3 参数更新

常用 SGD 更新规则：

$$W^{(l)} \leftarrow W^{(l)} - \eta \frac{\partial L}{\partial W^{(l)}} \quad (50)$$

$$b^{(l)} \leftarrow b^{(l)} - \eta \frac{\partial L}{\partial b^{(l)}} \quad (51)$$