# Multimodal Model Homework 1

March 30, 2025

**2023141460251 孙浩翔**

---

## 1 Question1: 向量是什么？在机器学习中向量可以用来表示什么？

**向量（Vector）**是数学中的基本概念，具有大小和方向的量。在代数中，向量通常表示为有序的数值列表：

$$\mathbf{v} = [v_1, v_2, \dots, v_n]$$

其中 $v_1, v_2, \dots, v_n$ 是向量的分量，$n$ 是向量的维度。

### 1.1 应用场景

- **特征表示**:
  数据通常以向量形式表示，如图像（像素值）、文本（词向量或 TF-IDF）、时间序列（数值向量）。
- **权重参数**:
  模型参数常以向量或矩阵形式表示，例如线性回归中的权重向量 $\mathbf{w}$ 或神经网络中的权重矩阵。
- **嵌入表示**:
  将离散对象映射到连续向量空间，例如 Word2Vec 的词嵌入、用户行为数据的低维表示。
- **距离与相似度**:
  计算向量之间的距离或相似度，用于分类、聚类等任务（如欧几里得距离、余弦相似度）。
- **优化问题**:
  参数更新通过向量实现，例如梯度下降法中对权重向量的迭代更新。

---

# 2 Question2: 如何度量不同向量的相似度？各自的优势如何？

## 2.1 常见的相似度度量方法

1. **欧几里得距离**
   表示两个向量之间的直线距离：
   $$d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}$$
   - **优势**: 直观易理解，适用于低维空间中的几何距离计算。
2. **余弦相似度**
   衡量两个向量的方向相似性：
   $$\text{cosine}(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$$
   - **优势**: 对向量长度不敏感，适用于高维稀疏数据（如文本分类、推荐系统）。
3. **曼哈顿距离**
   表示两个向量在各维度上的绝对差之和：
   $$d(\mathbf{u}, \mathbf{v}) = |u_1 - v_1| + |u_2 - v_2| + \cdots + |u_n - v_n|$$
   - **优势**: 计算简单，适用于网格状数据（如城市街区距离、路径规划）。
4. **Jaccard 相似度**
   衡量两个集合的交集与并集的比例：
   $$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$
   - **优势**: 适用于集合数据（如推荐系统中的用户兴趣匹配、文档相似性分析）。

---

# 3 Question3: 超平面是什么？它可以用来做什么？

**超平面（Hyperplane）**是一个 $n$ 维空间中的 $(n-1)$ 维子空间。
- 在二维空间中是直线，在三维空间中是平面。

## 3.1 应用场景

- **分类任务**: 支持向量机（SVM）使用超平面将数据分为两类。

- **回归分析**: 线性回归模型的预测值可以看作超平面上的点。

- **降维**: 超平面可用于主成分分析（PCA）等降维算法。

---

# 4 Question4: 什么是矩阵？你觉得矩阵在数据分析中的作用是什么？

**矩阵**（**Matrix**）是一个二维数组，用于表示多维数据或线性变换。

## 4.1 应用场景

- **数据表示**: 矩阵常用于存储表格数据（如特征矩阵）。
- **线性变换**: 矩阵可以表示旋转、缩放等几何变换。
- **优化问题**: 矩阵分解（如 SVD、PCA）用于降维和特征提取。
- **图结构**: 邻接矩阵表示图的连接关系。

---

# 5 Question5: 什么是线性相关与线性无关？

- **线性相关**: 若一组向量中存在一个向量可由其他向量的线性组合表示，则称这组向量线性相关。
- **线性无关**: 若一组向量中没有任何向量可由其他向量的线性组合表示，则称这组向量线性无关。

## 5.1 物理意义

线性无关的向量构成向量空间的一组基，能够唯一地表示该空间中的任意点。

---

# 6 Question6: 矩阵的秩描述了什么 (即它的物理意义)?

- **定义**: 矩阵的秩是其列向量或行向量的最大线性无关组的数量。
- **物理意义**:
  - 描述矩阵的“信息量”或“自由度”。
  - 决定线性方程组解的存在性和唯一性：若矩阵的秩等于未知数个数，则方程组有唯一解。

---

# 7 Question7: 伪逆如何计算？为什么我们需要伪逆？

## 7.1 伪逆

- **计算**: 使用奇异值分解（SVD）计算伪逆：

$$A^+ = V\Sigma^+U^T$$

  其中 $\Sigma^+$ 是对角矩阵 $\Sigma$ 的非零元素取倒数后转置得到的。
- **用途**:
  - 用于求解欠定或超定线性方程组。
  - 在最小二乘法中寻找最优解。

---

# 8 Question8: 流形是什么？举几个现实生活中的流形的例子，并给出流形学习的定义。

## 8.1 流形

- **定义**: 流形是一个局部类似于欧几里得空间的拓扑空间。
- **例子**:
  - 地球表面：二维流形嵌入在三维空间中。
  - 手写数字图像的空间：高维数据可能分布在低维流形上。

## 8.2 流形学习

- **定义**: 流形学习是一种降维技术，旨在发现高维数据中的低维流形结构。
- **应用**: 用于数据可视化（如 t-SNE）、特征提取（如 Isomap）。

---

# 9 Question9: 特征分解与奇异值分解的联系与不同。

## 9.1 联系

- 都是矩阵分解方法，用于提取矩阵的主要特征。

- 可用于降维、特征选择等任务。

### 9.2 不同

- **特征分解**: 仅适用于方阵，分解为特征值和特征向量：

$$A = Q\Lambda Q^{-1}$$

- **奇异值分解（SVD）**：适用于任意矩阵，分解为三个矩阵 $U$, $\Sigma$, $V^T$：

$$A = U\Sigma V^T$$

---

## 10 Question10: F 范数怎么计算？它可以用来度量什么？

### 10.1 计算公式

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$$

### 10.2 应用

- 衡量矩阵的整体大小。

- 在优化问题中作为正则化项，防止过拟合。

---

## 11 Question11: 给出微分的几何解释。

**微分**是函数在某一点的变化率，几何上表示为曲线的切线斜率。
- 微分反映了函数在局部的线性近似能力。

---

## 12 Question12: 梯度是如何计算的？

梯度是多元函数的偏导数组成的向量：

$$\nabla f(x_1, x_2, \dots, x_n) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right]$$

- **作用**: 梯度指向函数增长最快的方向，用于优化算法（如梯度下降）。

## 13 Question13: 局部最小值与全局最小值的关系。

- **局部最小值**: 在某个邻域内函数值最小。
- **全局最小值**: 在整个定义域内函数值最小。
- **关系**: 全局最小值一定是局部最小值，但局部最小值不一定是全局最小值。

---

## 14 Question14: 什么是凸函数、凸优化？

### 14.1 凸函数

- **定义**: 函数满足任意两点之间的连线在函数上方。
- **性质**: 凸函数的局部最小值即为全局最小值。

### 14.2 凸优化

- **定义**: 在凸函数上的优化问题，确保找到的局部最优解是全局最优解。
- **应用**: 广泛应用于机器学习（如线性回归、支持向量机）。

---

## 15 Question15: 什么是相关性、熵、KL 散度、交叉熵、最大似然？给出它们的定义及公式。

### 15.1 相关性

- **公式**:

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

  - 衡量两个变量的线性相关程度。