降维习题

浙江大学

赵洲

题目

- 1. **主成分分析（PCA）降维的主要目标是：**

- A. 增加数据的维度
- B. 保持数据的主要方差方向
- C. 减少数据的噪声
- D. 提高数据的非线性可分性

解答

- **A. 增加数据的维度**

- **错误原因**：PCA的主要目的是**减少**数据的维度，而不是增加。通过选择最具代表性的主成分，PCA在保持数据主要特征的同时降低了数据的复杂性。

解答

- **B. 保持数据的主要方差方向**

- **正确**：PCA通过线性变换找到数据中方差最大的方向，这些方向被称为主成分。保留这些主成分能够最大限度地保留数据的变异性和信息量。

解答

- **C. 减少数据的噪声**

- **部分正确**：虽然降维可能间接减少噪声，因为噪声通常分布在较低方差的方向，但PCA的主要目标不是专门去除噪声，而是保留数据的主要变异性。

解答

- **D. 提高数据的非线性可分性**

- **错误原因**：PCA是一种线性降维方法，不涉及非线性变换。因此，它并不能直接提高数据的非线性可分性。对于非线性可分性，通常需要使用非线性降维方法如t-SNE或Isomap。

题目

- **2. 在PCA中，选择前k个主成分时，通常是基于：**

- A. 主成分的方差贡献率
- B. 主成分的计算复杂度
- C. 主成分的相关性
- D. 主成分的稀疏性

解答

- **A. 主成分的方差贡献率**

- **正确**：在PCA中，选择前k个主成分通常基于它们的方差贡献率，即选择那些能够解释数据最大方差的主成分。这确保了降维后数据的主要信息被保留。

解答

- **B. 主成分的计算复杂度**

- **错误原因**：选择主成分时，计算复杂度不是主要考虑因素。PCA的目标是保留方差信息，而不是减少计算复杂度。

解答

- **C. 主成分的相关性**

- **错误原因**：PCA实际上使得主成分之间**不相关**。选择主成分的标准是它们的方差，而不是相关性。

解答

- **D. 主成分的稀疏性**

• **错误原因**：PCA并不关注主成分的稀疏性。稀疏性通常与其他降维方法（如稀疏PCA）相关，而非标准PCA。

题目

- **3. 在t-SNE算法中，主要用于降低维度的距离度量方法是:**

- A. 欧氏距离
- B. 曼哈顿距离
- C. Kullback-Leibler散度
- D. 余弦相似度

解答

- **A. 欧氏距离**

• **错误原因**：虽然t-SNE在高维空间中使用欧氏距离来计算相似度，但在降维过程中，其优化目标是最小化高维和低维空间之间的Kullback-Leibler散度。因此，欧氏距离不是t-SNE降维过程中的主要度量方法。

解答

- **B. 曼哈顿距离**

- **错误原因**：t-SNE不使用曼哈顿距离作为其主要的距离度量方法。曼哈顿距离在某些场景下可能适用，但不是t-SNE的核心距离度量。

解答

- **C. Kullback-Leibler散度**

- **正确**：t-SNE通过最小化高维空间和低维空间之间的Kullback-Leibler散度来优化低维表示。这种散度度量用于衡量两个概率分布之间的差异，是t-SNE降维的核心。

解答

- **D. 余弦相似度**

- **错误原因**：余弦相似度主要用于衡量向量间的角度相似性，而t-SNE并不使用它作为主要的距离度量方法。t-SNE更关注概率分布之间的差异。

题目

■ **4. 以下哪种降维方法是线性的?**

- A. t-SNE
- B. PCA
- C. Isomap
- D. LLE

解答

- **A. t-SNE**

- **错误原因**：t-SNE是一种**非线性**降维方法，主要用于高维数据的可视化，通过保留局部结构来展示数据的低维表示。

解答

- **B. PCA**

- **正确**：PCA是一种**线性**降维方法，通过线性组合原始特征来找到新的主成分方向，最大化数据在这些方向上的方差。

解答

- **C. Isomap**

- **错误原因**：Isomap是一种**非线性**降维方法，通过保持数据的测地距离（沿流形的距离）来实现降维。

解答

- **D. LLE**

- **错误原因**：局部线性嵌入（LLE）也是一种**非线性**降维方法，通过保持局部邻域的线性关系来实现降维。

题目

- 判断1.
- LDA（线性判别分析）不仅可以用于降维，还可以用于分类。

解答

- 正确

- LDA既是一种**降维**技术，通过寻找能够最大化类间方差和最小化类内方差的投影方向，也是一种**分类**方法，特别是在监督学习中，用于构建判别函数来区分不同类别。

题目

- 判断2.
- PCA在降维过程中会考虑数据的类别信息。

解答

- 错误

- **错误原因**：PCA是一种**无监督**的降维方法，它只关注数据的整体方差，**不**考虑任何类别信息或标签。PCA通过线性变换保留数据的主要变异性，而不区分类别，因此无法利用类别信息来指导降维过程。

题目

- 判断3.
- 降维后，数据的解释性通常会提高。

解答

- 错误

- 降维的主要目的是减少数据的复杂性和维度，而不一定提高数据的解释性。实际上，降维可能会丢失一些原始数据的信息，使得某些特征难以解释。然而，降维可以通过简化数据结构，使得某些模式更易于识别，但这不等同于整体解释性的提高。

题目

- 判断4.
- t-SNE适合用于高维数据的可视化，尤其是二维或三维。

解答

- 正确

- t-SNE是一种**非线性**降维技术，特别适用于将高维数据降到**二维或三维**，以便于可视化。它通过保留数据点之间的局部结构，使得在低维空间中能够直观地展示数据的聚类和分布。

题目

- 1. 主成分分析（PCA）的计算

a) 计算数据的均值向量。

b) 对数据进行中心化。

c) 计算协方差矩阵。

d) 求出协方差矩阵的特征值和特征向量。

e) 选择第一个主成分，并将数据降维到一维。

样本	x_1	x_2
1	2	0
2	0	2
3	2	2
4	0	0

解答

a) 计算均值向量

$$\mu = \left(\frac{2+0+2+0}{4}, \frac{0+2+2+0}{4}\right) = (1,1)$$

详细解释：

- 对每个特征（x_1和x_2）分别求平均值。
- 计算过程：
  - x_1的平均值：$(2+0+2+0)/4 = 1$
  - x_2的平均值：$(0+2+2+0)/4 = 1$
- 因此，均值向量为$\mu = (1,1)$。

解答

b) 对数据进行中心化

$$
\begin{aligned}
&\text{样本1}: (2-1, 0-1) = (1, -1)\\
&\text{样本2}: (0-1, 2-1) = (-1, 1)\\
&\text{样本3}: (2-1, 2-1) = (1, 1)\\
&\text{样本4}: (0-1, 0-1) = (-1, -1)
\end{aligned}
$$

详细解释：

- 中心化是指从每个数据点中减去均值向量，以使数据的均值为零。
- 计算过程：
  - 样本1：$(2-1, 0-1) = (1, -1)$
  - 样本2：$(0-1, 2-1) = (-1, 1)$
  - 样本3：$(2-1, 2-1) = (1, 1)$
  - 样本4：$(0-1, 0-1) = (-1, -1)$
- 中心化后的数据集为：$[(1, -1), (-1, 1), (1, 1), (-1, -1)]$

解答

c) 计算协方差矩阵

$$\text{Cov} = \frac{1}{4}\begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

详细解释：

- 协方差矩阵的计算公式为：

$$\text{Cov}(X) = \frac{1}{n}X^T X$$

解答

其中，X是中心化后的数据矩阵，n是样本数量。

- 中心化后的数据矩阵：

$$X = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix}$$

- 计算$X^T X$：

$$X^T X = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

- 因此，协方差矩阵为：

$$\mathrm{Cov} = \frac{1}{4} \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

解答

d) 求协方差矩阵的特征值和特征向量

协方差矩阵为单位矩阵，其特征值和特征向量如下：

$$\lambda_1 = 1, \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\lambda_2 = 1, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

详细解释：

- 单位矩阵的特征值均为1，对应的特征向量可以是标准基向量 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$。
- 由于协方差矩阵是对角矩阵，各对角线元素即为特征值，且特征向量是标准基向量。

解答

e) 选择第一个主成分，并将数据降维到一维

选择$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$作为主成分。

投影计算:

$$\begin{aligned}
&\text{样本}1: 1 \times 1 + (-1) \times 0 = 1 \\
&\text{样本}2: -1 \times 1 + 1 \times 0 = -1 \\
&\text{样本}3: 1 \times 1 + 1 \times 0 = 1 \\
&\text{样本}4: -1 \times 1 + (-1) \times 0 = -1
\end{aligned}$$

降维后的一维数据：1, -1, 1, -1。

解答

详细解释：

- 主成分$\mathbf{v}_1$为$\begin{bmatrix}1\\0\end{bmatrix}$，表示沿$x_1$轴的方向。
- 投影过程为将每个中心化后的数据点与主成分进行点积计算，得到一维表示。
  - 样本1：$1 \times 1 + (-1) \times 0 = 1$
  - 样本2：$-1 \times 1 + 1 \times 0 = -1$
  - 样本3：$1 \times 1 + 1 \times 0 = 1$
  - 样本4：$-1 \times 1 + (-1) \times 0 = -1$
- 降维后的数据为$[1, -1, 1, -1]$。

总结： 通过PCA，原始二维数据被投影到一维主成分上，保留了数据在x_1方向上的变异性。

题目

- **2. LDA的计算**
- 假设有两类数据，各有两个样本
- 使用线性判别分析（LDA）进行降维，求投影方向。

类1：(2,0), (4,0)

类2：(0,2), (0,4)

解答

1. **计算每类的均值：**

$$\mu_1 = \left(\frac{2+4}{2}, \frac{0+0}{2}\right) = (3,0)$$

$$\mu_2 = \left(\frac{0+0}{2}, \frac{2+4}{2}\right) = (0,3)$$

详细解释：

- 类1的均值：
  - x_1均值：$(2+4)/2=3$
  - x_2均值：$(0+0)/2=0$
- 类2的均值：
  - x_1均值：$(0+0)/2=0$
  - x_2均值：$(2+4)/2=3$

解答

2. **计算类内散布矩阵S_W：**

- 对于类1：

$$S_{W1} = \sum_{i=1}^{2}(\mathbf{x}_i - \mu_1)(\mathbf{x}_i - \mu_1)^T = \begin{bmatrix}(2-3)^2 & 0 \\ 0 & (0-0)^2\end{bmatrix} + \begin{bmatrix}(4-3)^2 & 0 \\ 0 & (0-0)^2\end{bmatrix} = \begin{bmatrix}1 & 0 \\ 0 & 0\end{bmatrix} + \begin{bmatrix}1 & 0 \\ 0 & 0\end{bmatrix} = \begin{bmatrix}2 & 0 \\ 0 & 0\end{bmatrix}$$

- 对于类2：

$$S_{W2} = \sum_{i=1}^{2}(\mathbf{x}_i - \mu_2)(\mathbf{x}_i - \mu_2)^T = \begin{bmatrix}(0-0)^2 & 0 \\ 0 & (2-3)^2\end{bmatrix} + \begin{bmatrix}(0-0)^2 & 0 \\ 0 & (4-3)^2\end{bmatrix} = \begin{bmatrix}0 & 0 \\ 0 & 1\end{bmatrix} + \begin{bmatrix}0 & 0 \\ 0 & 1\end{bmatrix} = \begin{bmatrix}0 & 0 \\ 0 & 2\end{bmatrix}$$

- 总的类内散布矩阵：

$$S_W = S_{W1} + S_{W2} = \begin{bmatrix}2 & 0 \\ 0 & 2\end{bmatrix}$$

解答

详细解释：

- 类内散布矩阵衡量每个类内数据的分散程度。
- 对于每个类，计算每个样本与该类均值的差，并计算其外积，然后累加。
- 类1的样本偏离均值的差分别为$(2-3, 0-0) = (-1, 0)$和$(4-3, 0-0) = (1, 0)$，其散布矩阵为：

$$S_{W1} = (-1)^2 + (1)^2 = 2 \quad \text{在} x_1 \text{ 方向上有散布}$$

- 类2的样本偏离均值的差分别为$(0-0, 2-3) = (0, -1)$和$(0-0, 4-3) = (0, 1)$，其散布矩阵为：

$$S_{W2} = (-1)^2 + (1)^2 = 2 \quad \text{在} x_2 \text{ 方向上有散布}$$

- 总的类内散布矩阵为：

$$S_W = S_{W1} + S_{W2} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

解答

3. **计算类间散布矩阵S_B：**

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T = \begin{bmatrix} 3-0 \\ 0-3 \end{bmatrix} \begin{bmatrix} 3 & -3 \end{bmatrix} = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$$

详细解释：

- 类间散布矩阵衡量不同类之间均值的分离程度。
- 计算方法是类间均值差的外积。
- $\mu_1 - \mu_2 = (3-0, 0-3) = (3, -3)$
- 类间散布矩阵：

$$S_B = \begin{bmatrix} 3 \\ -3 \end{bmatrix} \begin{bmatrix} 3 & -3 \end{bmatrix} = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$$

解答

4. 计算$S_W^{-1}S_B$：

$$S_W^{-1}=\frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

$$S_W^{-1}S_B=\frac{1}{2}\begin{bmatrix}9 & -9\\ -9 & 9\end{bmatrix}=\begin{bmatrix}4.5 & -4.5\\ -4.5 & 4.5\end{bmatrix}$$

详细解释：

- 计算S_W^{-1}，由于S_W是对角矩阵，逆矩阵也容易计算：

$$S_W^{-1}=\frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

- 计算$S_W^{-1}S_B$：

$$S_W^{-1}S_B=\frac{1}{2}\times\begin{bmatrix}9 & -9\\ -9 & 9\end{bmatrix}=\begin{bmatrix}4.5 & -4.5\\ -4.5 & 4.5\end{bmatrix}$$

解答

5. **求特征值和特征向量：**

矩阵 $\begin{bmatrix} 4.5 & -4.5 \\ -4.5 & 4.5 \end{bmatrix}$ 的特征值和特征向量如下：

- **特征值：** 9 和 0
- **对应的特征向量：**
  - 对于特征值9：$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$
  - 对于特征值0：$\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

解答

详细解释：

- 求解特征值：

$$\det\left(\begin{bmatrix}4.5-\lambda & -4.5\\ -4.5 & 4.5-\lambda\end{bmatrix}\right)=(4.5-\lambda)^2-(-4.5)^2=0$$

$$(4.5-\lambda)^2=20.25 \implies 4.5-\lambda=\pm4.5 \implies \lambda=9\text{ 或 }0$$

- 对于$\lambda=9$：

$$\begin{bmatrix}4.5-9 & -4.5\\ -4.5 & 4.5-9\end{bmatrix}=\begin{bmatrix}-4.5 & -4.5\\ -4.5 & -4.5\end{bmatrix}$$

解方程：

$$-4.5x-4.5y=0 \implies x=-y$$

选择特征向量为$\begin{bmatrix}1\\ -1\end{bmatrix}$。

- 对于$\lambda=0$：

$$\begin{bmatrix}4.5 & -4.5\\ -4.5 & 4.5\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}=0$$

解方程：

$$4.5x-4.5y=0 \implies x=y$$

选择特征向量为$\begin{bmatrix}1\\ 1\end{bmatrix}$。

解答

6. **选择最大特征值对应的特征向量作为投影方向：**

$$\mathbf{w} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

详细解释：

- 在LDA中，选择对应于最大特征值的特征向量作为投影方向，以最大化类间散布与类内散布的比率。
- 最大特征值为9，对应的特征向量为$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$，因此投影方向为$(1, -1)$。

总结： 通过LDA，找到最能区分两类数据的投影方向为$(1, -1)$，即沿此方向将数据降维。

题目

- 3. 奇异值分解（SVD）与PCA
- 解释奇异值分解（SVD）如何用于实现PCA，并说明其中的关键步骤。

解答

奇异值分解（SVD）可以用于计算PCA的主成分。关键步骤如下：

1. **数据中心化**：将数据矩阵X的每个特征减去其均值，使数据中心化。

   详细解释：

   - 数据中心化是PCA的第一步，确保每个特征的均值为零，以避免均值对主成分的影响。
   - 中心化后的数据矩阵X具有零均值。

解答

2. **SVD分解**：对中心化后的数据矩阵X进行SVD分解：

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

其中，U和V是正交矩阵，Σ是对角矩阵，包含奇异值。

详细解释：

- SVD将数据矩阵分解为三个矩阵的乘积。
- U：表示数据在左奇异向量方向上的投影。
- Σ：包含奇异值，反映数据在各主成分方向上的方差。
- V^T：表示主成分方向，即特征向量。

解答

3. **主成分选择**：矩阵V的列向量即为PCA的主成分。选择前k个对应最大奇异值的列向量作为前k个主成分。

   详细解释：

   - 奇异值的大小与对应主成分的方差成正比。
   - 选择前k个最大奇异值对应的主成分，确保最大程度地保留数据的变异性。

解答

4. **数据投影**：将数据投影到选择的主成分上，即计算：

$$X_{\text{降维}} = XV_k$$

其中，V_k是前k个主成分组成的矩阵。

详细解释：

- 通过将中心化后的数据矩阵与主成分矩阵相乘，实现数据在新低维空间的表示。
- 结果$X_{\text{降维}}$是降维后的数据，保留了主要的方差信息。

解答

- **总结：** 通过SVD，PCA的计算过程变得更加高效，尤其适用于大规模数据集。SVD分解提供了主成分和对应的奇异值，帮助我们选择保留最重要的特征进行降维。

End