2025 模式识别
作业四

人工智能学院 221300079 王俊童

2025.4.15

221300079 王俊童，人工智能学院

1 问题一

a. 我们假设对于 i 样本，K 近邻 $z_{i1}, \cdots, z_{iK}$, 设 $Z_i = [z_{i1} - z_i, \cdots, z_{iK} - z_i]$, 有重构误差:

$$\|\sum_{j=1}^{K} w_{ij}(z_{ij} - z_i)\|^2 = w_i^T C_i w_i = w_i^T Z_i^T Z_i w_i$$

所以优化问题如下：

$$\min_{w_i} \quad w_i^T C_i w_i$$
$$s.t. \sum_{j=1}^{K} w_{ij} = 1$$

根据拉格朗日乘子法:

$$L(w_i, \lambda) = w_i^T C_i w_i - \lambda(\mathbf{1}^T w_i - 1), \quad C_i w_i = \frac{\lambda}{2}\mathbf{1}, \quad \mathbf{1}^T w_i = 1$$

解得:

$$\lambda = \frac{2}{\mathbf{1}^T C_i^{-1}\mathbf{1}}, \quad w_i = \frac{C_i^{-1}\mathbf{1}}{\mathbf{1}^T C_i^{-1}\mathbf{1}}$$

b. 根据第一问，我们逐个进行证明:

- 旋转不变性，令 $z_j' = Qz_j$, 因为 Q 是正交矩阵 $Q^TQ = I$:

$$\|Qz_i - \sum_j w_{ij} Q z_j\|^2 = \|Q(z_i - \sum_j w_{ij} z_j)\|^2 = \|(z_i - \sum_j w_{ij} z_j)\|^2$$

- 平移不变性, 令 $z_j' = z_j + t$:

$$\|z_i + t - \sum_j w_{ij}(z_j + t)\|^2 = \|z_i - \sum_j w_{ij} z_j + t - (\sum_j w_i j)t\| = \|(z_i - \sum_j w_{ij} z_j)\|^2$$

- 缩放不变性, 令 $z_j' = sz_j$:

$$\|sz_i - \sum_j w_{ij} s z_j\|^2 = \|s^2(z_i - \sum_j w_{ij} z_j)\|^2 \to \|(z_i - \sum_j w_{ij} z_j)\|^2$$

其中缩放倍数 s 并不改变优化目标。

c. 首先这么修改之后，因为这个问题本质是一个 manifold learning 的问题。优化目标不改变，仍然是在一个低维空间里面重构一个权重，保留了局部的 manifold 的结构。

对于两个约束条件，第一个等于 0 的约束相当于约束了所有低维向量的均值为 0，消除了平移自由度，这样低维表示都在原点附近了。

对于第二个条件，相当于正交而且归一化，这样协方差就是单位了，消除了缩放，而且整个低维都在一个单位球上，所以表示也被规范化了。所以好像旋转自由度也被消除了。

其实我个人还有一个理解是，这个问题对于旋转自由度其实不是特别的敏感，因为好像旋转自由度对于这个问题影响真的不大，似乎最优解都是在一个点附近了。

d. 不妨由厦门政府，我们从 9.52 开始算：

$$L(y) = \sum_{i=1}^{n} \|y_i - \sum_j w_{ij} y_j\|^2 = \|Y - WY\|^2$$

这个问题是等价的事，我们可以令 $\hat{y} = \sum_j w_{ij} y_j$, 然后写成矩阵形式. 可以化简

$$L(y) = Tr(Y^T (I - W)^T (I - W) Y) = Tr(Y^T M Y) = \sum_{i,j} M_{ij} y_i^T y_j$$

e. 首先证明 M 是一个半正定的, 假设有 v:

$$v^T M v = \|(I - W)v\|^2 \geq 0$$

然后为什么 1 向量是一个特征向量呢:

$$(I - W)\mathbf{1} = 1 - W1 = \mathbf{0}, \quad \sum_j w_{ij} = 1$$

所以证明完了，然后这个第一个按照升序排序的就是 0 对应的特征向量。因为我们在降维的时候实际上对于的是 d 个最小的非零特征值的特征向量。所以这个对于平移自由度的约束其实是没用的，所以要舍弃，无意义。

f. 尝试看看但是:

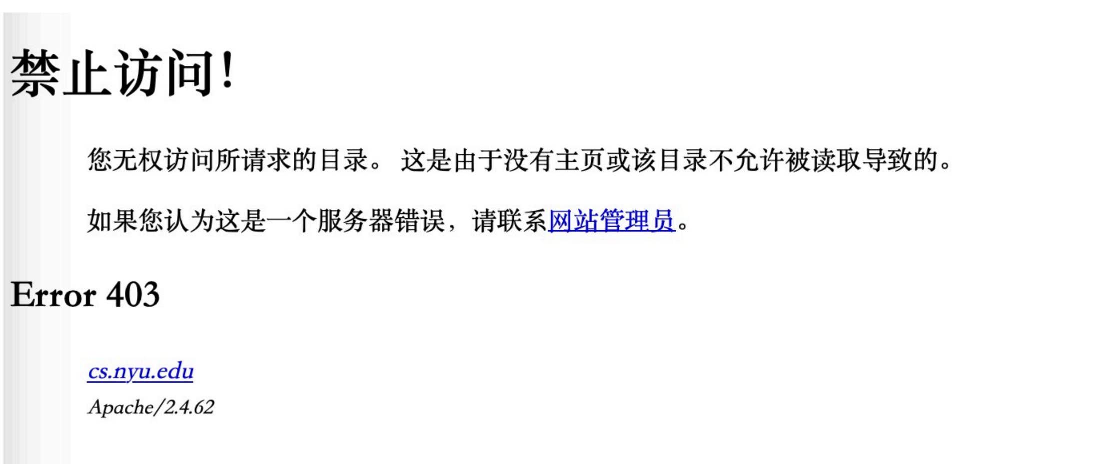

Figure 1: forbidden

2 问题二

a.

$$1-\sigma(x)=1-\frac{1}{1+e^{-x}}=\frac{e^{-x}}{1+e^{-x}}=\frac{1}{1+e^{x}}=\sigma(-x)$$

b.

$$\sigma(x)'=\frac{e^{-x}}{(1+e^{-x})^2}=\sigma(x)(1-\sigma(x))$$

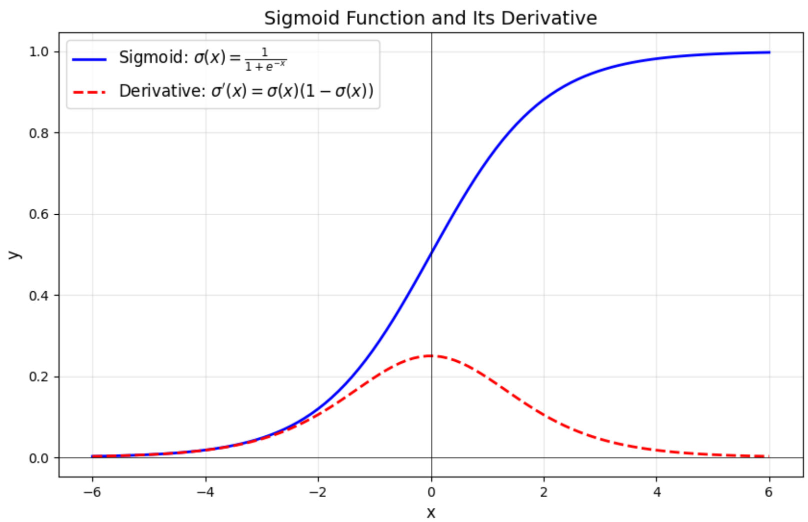

Figure 2: sigmoid

c. 因为神经网络是一个多层的，并且是一层一层求导往后算的，所以我们目前看第 i 层的网络节后

$$y_j^{(i)}=\sigma'(z_j^{(i)})$$

所以根据上面的图可以看到，这个导数是一个只有最大值 0.25 的情况，并且越往旁边数据就越接近 0，有大一点的值的位置是非常少的。所以我们如果从头开始算 L 次，梯度会变得非常非常少，这就会导致深度神经网络中的梯度消失问题。结果越趋近于 0，问题越大。

3 问题三

a. 1, 1, 2, 3, 5, 8

b. 可以求解 $F_n-F_{n-1}-F_{n-2}=0$, 我们假设通解如下:

$$F_n=A\alpha^n+B\beta^n$$

所以根据特征方程可以得到:

$$r^2-r-1=0,\quad \alpha=\frac{1+\sqrt{5}}{2},\beta=\frac{1-\sqrt{5}}{2}$$

因为根据前两个斐波那契数可以得到：

$$\begin{cases} A\alpha + B\beta = 1 \\ A\alpha^2 + B\beta^2 = 1 \end{cases}$$

可以解答得到：$A = \frac{1}{\alpha-\beta}, B = \frac{-1}{\alpha-\beta}$, 所以得到:

$$F_n = \frac{\alpha^n - \beta^n}{\alpha - \beta}$$

c. 我们首先证明这个式子，用数学归纳法；

- 首先为 1 的情况：$F_3 = 2$, $F_1 = 1$, 所以满足。
- 假设 n 也满足
- 看 n+1 的情况:

$$\sum_{i=1}^{n+1} F_i = (\sum_{i=1}^{n} F_i) + F_{n+1} = F_{n+2} - 1 + F_{n+1} = F_{n+3} - 1$$

然后我们看

$$p_i = \frac{F_i}{F_{n+2} - 1}, i = 1, 2, \cdots, n$$

首先这个东西肯定非负，然后在 0-1 之间，然后还满足:

$$\sum p_i = \frac{\sum F_i}{F_{n+2} - 1} = 1$$

d. 用数学归纳法证明:

- n=1, $\sum 1F_1 = F_3 - F_4 + 2 = 1$
- n=k，假设也满足这个式子
- n=k+1: 要证明

$$\sum_{i=1}^{k+1} iF_i = (k+1)F_{k+3} - F_{k+4} + 2$$

$$\sum_{i=1}^{k+1} iF_i = \sum_{i=1}^{k} iF_i + (k+1)F_{k+1} = kF_{k+2} - F_{k+3} + 2 + (k+1)F_{k+1}$$

$$= (k+1)(F_{k+1} + F_{k+2}) - (F_{k+2} + F_{k+3}) + 2 = (k+1)F_{k+3} - F_{k+4} + 2$$

所以证明完毕

e. 首先这个 p 为 $F_1, \cdots, F_5 = \{\frac{1}{12}, \frac{1}{12}, \frac{2}{12}, \frac{3}{12}, \frac{5}{12}\}$ 所以霍夫曼树如下:

```
        [12]
       /    \
    [5]      [7]
    (E)     /   \
        [3]      [4]
        (D)     /   \
            [2]      [2]
            (C)     /   \
                 (A)    (B)
```

其中：$E : 0, D : 10, C : 110, A : 1110, B : 1111.$

f. 根据观察，我们可以看到由于斐波那契数列的单调性质，从 n=2 开始，就满足每一个都比上一个大，所以树一定是不平衡的，意思是，我们的 $F_1, ..., F_n$ 对应的编码长度是:$n-1, n-1, n-2, ..., 1$，所以满足:

$$B = p * l = \frac{1}{F_{n+2}-1} * E, \quad E = (\sum_{i=2}^{n} F_i(n-i+1) + (n-1))$$

因为:

$$\sum_{i=2}^{n} F_i n = F_{n+2} - 2, \quad \sum_{i=2}^{n} iF_i = nF_{n+2} - F_{n+3} + 1$$

所以可以化简为:

$$E = n(F_{n+2}-2) - (nF_{n+2} - F_{n+3} + 1) + (F_{n+2}-2) + n - 1 = F_{n+3} + F_{n+2} - n - 4 = F_{n+4} - (n+4)$$

$$B = \frac{F_{n+4} - (n+4)}{F_{n+2} - 1}$$

g. 因为 $F_n = \frac{\alpha^n - \beta^n}{\sqrt{5}}$ 因为有一个 $\frac{1-\sqrt{5}}{2} < 0$, 所以 $n \to \infty, \quad F_n \sim \frac{\alpha^n}{\sqrt{5}}$

所以:

$$B_n \approx \frac{\frac{\alpha^{n+4}}{\sqrt{5}} - (n+4)}{\frac{\alpha^{n+2}}{\sqrt{5}} - 1}$$

可以忽略低阶项:

$$B_n \approx \frac{\frac{\alpha^{n+4}}{\sqrt{5}}}{\frac{\alpha^{n+2}}{\sqrt{5}}} = (\frac{1+\sqrt{5}}{2})^2 \approx 2.618$$

所以平均需要 2.618 的比特来编码。

4 问题四

a. 根据积分图的定义可以得到 B:

$$\begin{pmatrix} 1 & 1 & -1 \\ 2 & 3 & 1 \\ 3 & 4 & 6 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 \\ 3 & 7 & 7 \\ 6 & 14 & 20 \end{pmatrix}$$

b. 我们根据积分图的定义，确实可以有一种动态规划的算法:

$$B(i,j) = A(i,j) + B(i-1,j) + B(i,j-1) - B(i-1,j-1)$$

根据动态规划算法来看，整个算法确实是 $O(m*n)$ 的，因为每一个位置的计算都是 $O(1)$ 的，然后确实是线性的。其实是两个大矩形相加减去重复部分就行。

c. 那么根据之前我们的积分图来说，任意的矩阵 $i_1 i_2 j_1 j_2$ 都可以 $O(1)$ 算出来，我们只需要做 3 次加减法即可，相当于:

$$S = B(i_2,j_2) - B(i_1-1,j_2) - B(i_2,j_1-1) + B(i_1-1,j_1-1)$$

所以是 $O(1)$ 的。

5 问题五

a. 我们只需要证明一个问题即可证明独立性：$p(A,B|C)=p(A|C)p(B|C)$:

$$p(A,B|C)=\frac{p(A,B,C)}{p(C)}=\frac{p(C|A)p(B|C)p(C)}{p(C)}=\frac{p(A,C)}{p(C)}p(B|C)=p(A|C)p(B|C)$$

b. 1.2 同理:

$$p(A,B|C)=\frac{p(A,B,C)}{p(C)}=\frac{p(B)p(C|B)p(A|C)}{p(C)}=\frac{p(B,C)}{p(C)}p(A|C)=p(A|C)p(B|C)$$

c. 1.3 同理:

$$p(A,B|C)=\frac{p(A,B,C)}{p(C)}=\frac{p(C)p(A|C)p(B|C)}{p(C)}=p(A|C)p(B|C)$$

d. 若 C 没有被观察到:

$$p(A,B)=\sum_C p(A,B,C)=\sum_C p(C|A,B)p(A)p(B)=p(A)p(B)$$

所以 A 和 B 独立，一个例子就是如果 $C=A\oplus B$，那如果 p 为 0.5 的时候，是独立的。但是一旦我们观测到比如 C=0 的时候，那就是一样的，1 就是不一样的。

e. 由于 F 是 C 的后代，不独立，所以有:

$$p(C|F)=\frac{p(C,F)}{p(F)}=p(C)\frac{p(F|C)}{p(F)}\neq p(C)$$

所以给定 F 的情况，然后我们就知道了 A 和 B 不再是独立的，而是依存的。