《计算机应用数学第三次作业》

提醒：本次作业计算题与编程题的截止时间不同，请分别提交至对应的窗口。

计算题（60 分）截止时间：2024.06.04 23:59

1. (15 分）对于给定矩阵 A（规模为 4×2），求 A 的 SVD（奇异值分解），即求U，Σ，V^{T}，使得$\mathrm{A} = \mathrm{U}\Sigma\mathrm{V}^{\mathrm{T}}$。其中$\mathrm{U}^{\mathrm{T}}\mathrm{U} = \mathrm{I}$，$\mathrm{V}^{\mathrm{T}}\mathrm{V} = \mathrm{I}$，给出求解过程。

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \\ -1 & 0 \end{bmatrix}$$

2. (15 分）现有如下数据（5 个样本，4 个维度（A，B，C，D）），用 PCA 将数据降到 2 维，给出求解过程。

A	B	C	D
1	5	3	1
4	-4	6	6
1	4	3	2
4	4	2	2
5	5	2	4

3. （15 分）假设有六个样本在二维空间中：
正样本：（1，1），（2，1），（2，3）
负样本：（-1，-1），（0，-3），（-2，-4）
使用支持向量机求划分平面，给出支持向量模型及求解过程。

4. （15 分）给定一个 4*3*2 的神经网络，权重矩阵为：

$$W_1 = \begin{bmatrix} 0.10 & 0.40 & 0.35 \\ 0.15 & 0.20 & 0.25 \\ 0.05 & 0.35 & 0.40 \\ 0.20 & 0.25 & 0.15 \end{bmatrix}$$

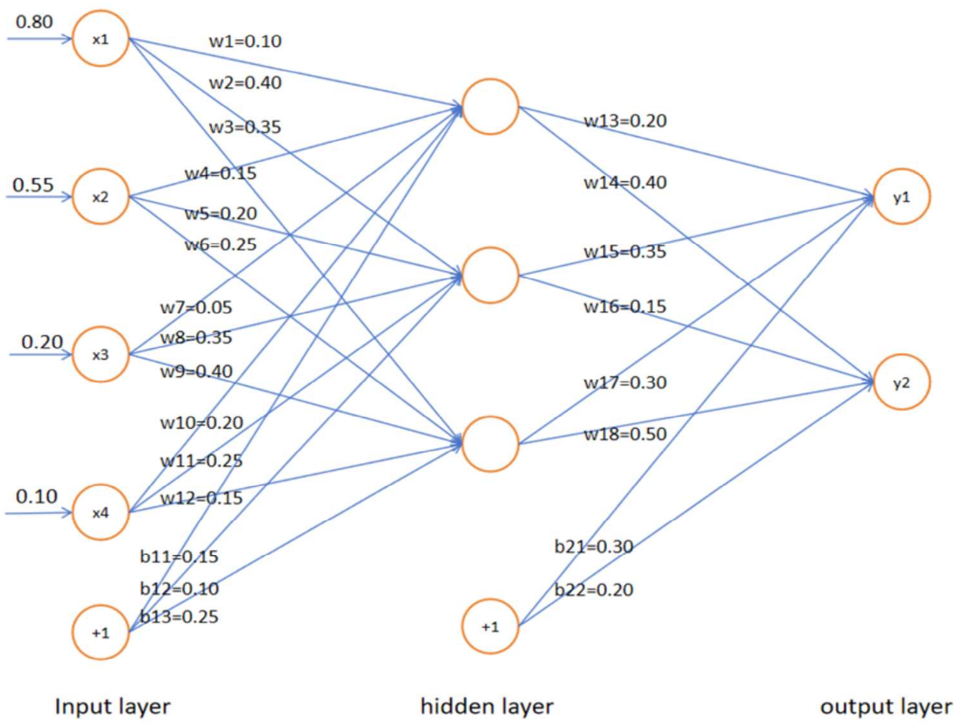
$$b_1 = [0.15,0.10,0.25]$$

$$W_2 = \begin{bmatrix} 0.20 & 0.40 \\ 0.35 & 0.15 \\ 0.30 & 0.50 \end{bmatrix}$$

$$b_2 = [0.30,0.20]$$

激活函数为 sigmoid 函数。

给定一个训练样本 x=[0.80,0.55,0.20,0.10]和 y=[1,0]，假设学习率为 0.01，损失函数为二元交叉熵损失，计算w_{10}，w_{13}参数的一次更新结果。



编程题 （40 分）截止时间：2024.06.06 23:59

1、基于降维的机器学习（20 分）

数据集：kddcup99_Train.zip, kddcup99_Test.zip

数据描述：https://www.kdd.org/kdd-cup/view/kdd-cup-1999/Data，目的是通过 42 维的数据判断某一个数据包是否为 attack。(注意，数据中有多种类型的 attack，我们将他们统一认为是 attack，不做具体区分，只区分 normal 和 attack，可以预处理的时候把所有的 attack 都统一 re-label，具体 attack 列表可参考 https://kdd.org/cupfiles/KDDCupData/1999/training_attack_types)。

任务描述：选择 SVM，结合降维方法 PCALDA，实现数据降维+分类。

要求输出：不同方法降低到不同维度对判别结果的影响（提升还是下降），Plot 出两种降维方法降到 3 维之后的结果（Normal 和 Attack 的 sample 用不同颜色）

2、深度学习训练方法总结（20 分）

数据集：kddcup99_Train.zip, kddcup99_Test.zip

任务描述：实现一个简单的神经网络模型判别数据包是否为 attack，网络层数不小于 5 层，平台不限（TensorFlow、Pytorch 等都可），例如 42->36->24->12->6->1。尝试至少 2 种

激活函数，至少 2 种 parameter initialization 方法，至少 2 种训练方法（SGD，SGD+Momentom，Adam），训练模型并判断训练结果
要求输出：1）模型描述，层数，每一层参数，激活函数选择，loss 函数设置等；2）针对不同方法组合（至少 2 个组合），plot 出随着 epoch 增长 training error 和 test error 的变化情况。

如何求奇异值分解矩阵

- 由于Σ除了对角线上是奇异值其他位置都是0，那只需要求出每个奇异值σ：

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

$$\Rightarrow Av_i = \sigma_i u_i \Rightarrow \sigma_i = Av_i / u_i$$

这样可以求出每个奇异值，进而求出奇异值矩阵Σ

$$A = U\Sigma V^T \Rightarrow A^T = V\Sigma U^T$$
$$\Rightarrow A^T A = V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T$$

特征值矩阵等于奇异值矩阵的平方，也就是说特征值和奇异值满足如下关系：$\sigma_i = \sqrt{\lambda_i}$

定矩阵 A（规模为 $m\times n$），求 A 的 SVD（奇异值分解），即求 U,Σ,V^T，使得 $A=U\Sigma V^T$，其

$U=I$, $V^TV=I$。

$$A=\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \\ -1 & 0 \end{bmatrix}$$

4×2 矩阵

$$A=\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \\ -1 & 0 \end{bmatrix}$$

$A = U\Sigma V^T$

$\underset{4\times4}{UU^T = I}$ $\underset{2\times2}{VV^T = I}$

$\Sigma = \mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2})$

$W = A^T A = (U\Sigma V^T)^T U\Sigma V^T = V\Sigma^2 V^T$

$(W-\lambda I)x = 0$

$\lambda_1 \geq \lambda_2 \geq 0$ $V = [v_1\ v_2]$

$AA^T = W'$

$(W'-\lambda I)x' = 0$

$\lambda_1' \geq \lambda_2' \geq \lambda_3' \geq \lambda_4' \geq 0$

$U = [u_1\ u_2\ u_3\ u_4]$

下数据（5 个样本，4 个维度（A、B、C、D）），用 PCA 将数据降到 2 维。

A	B	C	D
1	5	3	1
4	-4	6	6
1	4	3	2
4	4	2	2
5	5	2	4

PCA 算法

设有m条n维数据：

- 1）将原始数据按列组成n行m列矩阵X
- 2）将 X 的每一行（代表一个属性字段）进行零均值化，即减去这一行的均值
- 3）求出协方差矩阵$C = \frac{1}{m}XX^T$
- 4）求出协方差矩阵的特征值及对应的特征向量
- 5）将特征向量按对应特征值大小从上到下按行排列成矩阵，取前k行组成矩阵P
- 6）$Y = PX$即为降维到k维后的数据

$$X = \begin{bmatrix} 1 & 4 & 1 & 4 & 5 \\ 5 & -4 & 4 & 4 & 5 \\ 3 & 6 & 3 & 2 & 2 \\ 1 & 6 & 2 & 2 & 4 \end{bmatrix} \begin{matrix} 3 \\ 2.8 \\ 3.2 \\ 3 \end{matrix}$$

$X_{center} = X_c$

$\mathrm{cov}(X_c) \frac{1}{5} X_c X_c^T$

$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq \lambda_5 \geq 0$

$v_1 \quad v_2 \quad v_3 \quad v_4 \quad v_5$

$P = [v_1 \ v_2]$

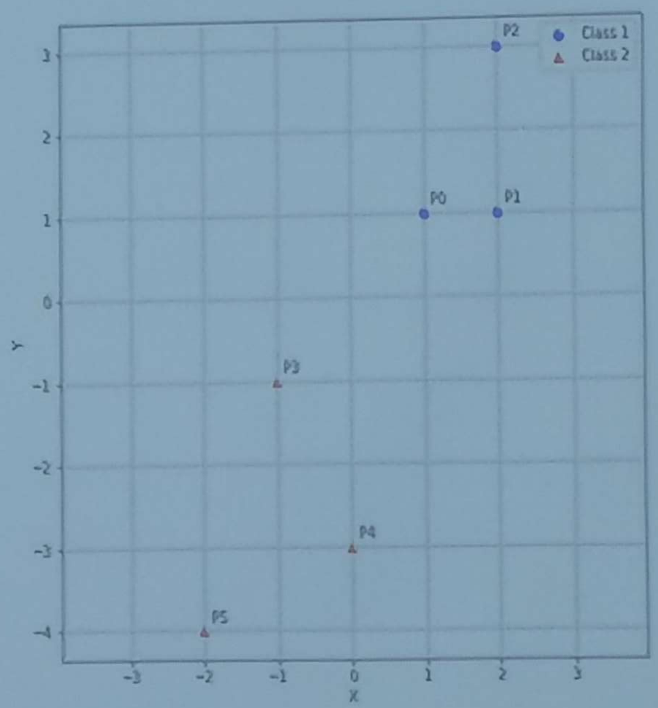
PX

3. (15 分) 假设有六个样本在二维空间中：

正样本：(1, 1)、(2, 1)、(2, 3)

负样本：(-1, -1)、(0, -3)、(-2, -4)

使用支持向量机求划分平面，给出支持向量模型及求解过程。



令划分的超平面为：$\boldsymbol{w^T x + b = 0}$

目标：找到一个超平面，使所有样本分类正确，且支持向量到超平面距离最远

$$\begin{cases} w^T x_l + b \le -\rho/2 \text{ if } y_l = -1 \\ w^T x_l + b \ge \rho/2 \text{ if } y_l = 1 \end{cases} \Leftrightarrow \quad y_l(w^T x_l + b) \ge \rho/2$$

放缩后得到：

$$\begin{cases} w^T x_l + b \le -1 \text{ if } y_l = -1 \\ w^T x_l + b \ge 1 \text{ if } y_l = 1 \end{cases} \Leftrightarrow \quad y_l(w^T x_l + b) \ge 1$$

$w^T x + b \Rightarrow$

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i (y_i(w x_i + b) - 1), \quad \alpha_i \ge 0$$

$$\theta(w) = \max_{\alpha_i \ge 0} L(w,b,\alpha)$$

$$L(w,b,\alpha) = \max_{\alpha} \left\{ \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i}^{m} \sum_{j} \alpha_i \alpha_j y_i y_j x_i^T x_j \right\}$$

在KKT条件下，原问题最优解为对偶问题的最优解

$$\min_{\alpha} \left\{ \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{m} \alpha_i \right\}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} \alpha_i y_i = 0$$

$$\alpha_i \ge 0$$

$$y_i(w^T x_i + b) - 1 \ge 0$$

$$a_i(y_i(w^T x_i + b) - 1) = 0$$

通过SMO算法求解最优 α：

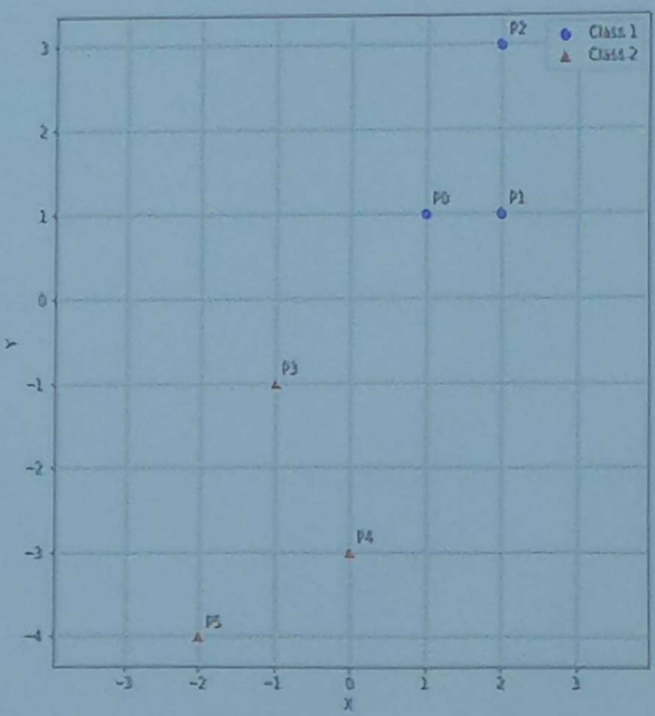
1.选择两个需要更新的参数α_i和α_j，固定其他参数。于是有以下约束：

$$\alpha_i y_i + \alpha_j y_j = c, \alpha_i \ge 0, \alpha_j \ge 0$$

其中$c = -\sum_{k \ne i,j} \alpha_k y_k$，由此可以得出$\alpha_j = \frac{c-\alpha_i y_i}{y_j}$，也就可以用$\alpha_i$的表达式代替$\alpha_j$。这样就相当于把目标问题转化成了仅有一个约束条件的最优化问题，仅有的约束是 $\alpha_i \ge 0$。

2. 对于仅有一个约束条件的最优化问题，可以在α_i上对优化目标求偏导，令导数为零，从而求出变量值$\alpha_{i_{new}}$，然后根据$\alpha_{i_{new}}$求出 $\alpha_{j_{new}}$。

3. 多次迭代直至收敛。



例：选择支持向量（1，1），（-1，-1）

$$w^T = (0.5, 0.5), b = 0$$

$x_1 = \begin{pmatrix}1\\1\end{pmatrix}$ $x_2 = \begin{pmatrix}-1\\-1\end{pmatrix}$

$\min_{\alpha} \frac{1}{2}\left[2\alpha_1^2 + 2\alpha_2^2 + 4\alpha_1\alpha_2\right] - \alpha_1 - \alpha_2$

$\alpha_1 - \alpha_2 = 0$ $\alpha_1 = \alpha_2$

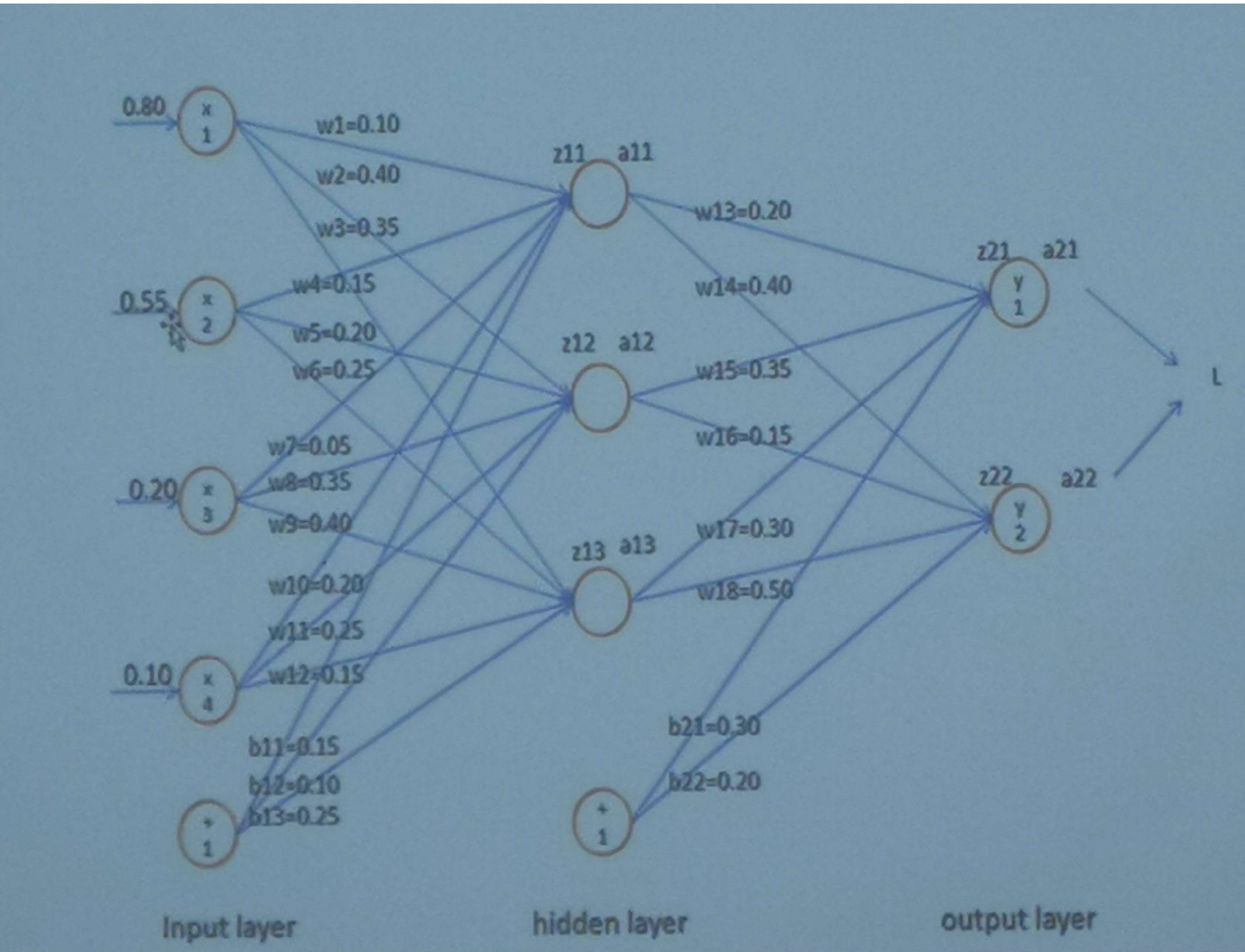
$2\alpha_1 - 4\alpha_1^2$

$2 - 8\alpha_1 = 0$

$\alpha_1 = \alpha_2 = \frac{1}{4}$

$w = \sum_i^m \alpha_i y_i x_i = \begin{pmatrix}0.5\\0.5\end{pmatrix}$

$x_1, w^T x + b = 1$ $b = 0$



$$L = -\frac{1}{2}\left[\log a_{21} + \log(1-a_{22})\right]$$

$$\frac{\partial L}{\partial w_{13}} = \frac{\partial L}{\partial a_{21}} \cdot \frac{\partial a_{21}}{\partial z_{21}} \cdot \frac{\partial z_{21}}{\partial w_{13}}$$

$$= -\frac{1}{2} \times \frac{1}{a_{21}} \times \sigma(z_{21}) \times (1-\sigma(z_{21}))\, a_{11}$$

$$= -0.0876$$

$$w_{13} = w_{13} - r \times \frac{\partial L}{\partial w_{13}} = 0.20876$$

$$\frac{\partial L}{\partial w_{10}} = \frac{\partial L}{\partial a_{21}} \frac{\partial a_{21}}{\partial z_{21}} \overset{w_{13}}{\frac{\partial z_{21}}{\partial a_{11}}} \frac{\partial a_{11}}{\partial z_{11}} \frac{\partial z_{11}}{\partial w_{10}}$$

$$-\frac{1}{2} \times \frac{1}{a_{21}} \times \sigma(z_{21})(1-\sigma(z_{21})) \times w_{13} \times \sigma(z_{11})(1-\sigma(z_{11})) \times x_4$$

$$+ \frac{\partial L}{\partial a_{22}} \frac{\partial a_{22}}{\partial z_{22}} \underset{w_{14}}{\frac{\partial z_{22}}{\partial a_{11}}} \frac{\partial a_{11}}{\partial z_{11}} \frac{\partial z_{11}}{\partial w_{10}} \times x_4$$

$$-\frac{1}{2} \times (-1) \times \frac{1}{1-a_{22}} \qquad \sigma(z_{11})(1-\sigma(z_{11}))$$

$$\sigma(z_{22})(1-\sigma(z_{22}))$$

$$= 0.002698$$

$$w_{10} = w_{10} - r \cdot \frac{\partial L}{\partial w_{10}} = 0.1997302$$