Problem 1. (15 分) 请证明泊松分布、均匀分布、正态分布的期望与方差。

Solution:

- 泊松分布：$E = \lambda, Var = \lambda$
- 均匀分布：$E = (a+b)/2, Var = (a-b)^2/12$
- 正态分布：$E = \mu, Var = \sigma^2$

Problem 2. 布隆过滤器 (Bloom Filter) 是一种空间效率高、查询效率快的数据结构，用于判断一个元素是否属于一个集合。在大规模数据的快速查询、去重等应用中使用广泛。它实质上是一个长度为 m 的 01 数组和 k 个不同的哈希函数构成。在添加元素时，将元素经过 k 个哈希函数得到多个哈希值，并将数组上这些哈希值位置标记为 1。在查询元素时，将元素经过同样的 k 个哈希函数得到多个哈希值，若数组上这些哈希值位置都为 1，则说明元素可能在集合中，否则一定不在集合中。布隆过滤器可能会出现误判，但不会出现漏判。

(a) (10 分) 假设每个哈希函数都是随机函数，请计算在插入 n 个元素后，布隆过滤器出现误判的概率，即一个不在集合中的元素被判定为在集合中。

(b) (15 分) 分析哈希函数的数量 k 对误判的影响，并计算使得误判概率最小的 k，假设 $n = C \cdot m$。

Solution:

(a) 假设 Hash 函数以等概率条件选择并设置数组中的某一位，那么位数组中某一特定的位在进行元素插入时的 Hash 操作中没有被置位的概率是：

$$1 - \frac{1}{m}. \tag{1}$$

那么在所有 k 次 Hash 操作后该位都没有被置 1 的概率是:

$$\left(1-\frac{1}{m}\right)^{k}. \tag{2}$$

如果我们插入了 n 个元素，那么某一位仍然为 0 的概率是:

$$\left(1-\frac{1}{m}\right)^{kn}. \tag{3}$$

因而该位为 1 的概率是:

$$1-\left(1-\frac{1}{m}\right)^{kn}. \tag{4}$$

现在检测某一元素是否在该集合中。标明某个元素是否在集合中所需的 k 个位置都按照如上的方法设置为 1，但是该方法可能会使算法错误的认为某一原本不在集合中的元素却被检测为在该集合中（False Positives），该概率由以下公式确定:

$$\left[1-\left(1-\frac{1}{m}\right)^{kn}\right]^{k} \approx \left(1-e^{-kn/m}\right)^{k}. \tag{5}$$

只需要答出公式 5 的左边部分就可以给满分。

(b) 对于公式 5 有

$$\left[1-\left(1-\frac{1}{n}\right)^{km}\right]^{k} = \left[1-\left(1-\frac{1}{n}\right)^{n\cdot\frac{km}{n}}\right]^{k} \approx \left[1-e^{-\frac{km}{n}}\right]^{k} = \left[1-e^{-\frac{k}{C}}\right]^{k}. \tag{6}$$

问题可以转换为最小化

$$f(k) = \left[1-e^{-\frac{k}{C}}\right]^{k}. \tag{7}$$

令 $x = k/C$，我们有

$$\begin{aligned} g(x) &= \ln f(k) = C\cdot x\ln\left(1-e^{-x}\right), \\ g^{'}(x) &= C\ln\left(1-e^{-x}\right) + Cx\cdot\frac{e^{-x}}{1-e^{-x}}. \end{aligned} \tag{8}$$

观察到对于 $x = \ln 2$ 有 $g^{'}(\ln 2) = 0$，并且在 $x > 0$ 时 $g^{'}(x)$ 单调增的。所以 $g(x)$ 在 $x = \ln 2$ 处最小。所以，使得误判概率最小的 $k = C\cdot\ln 2$。

Problem 3. (10 分) 我们有两种硬币：一种是公平的硬币，即抛一次正反的概率均为 1/2；另一种是产生正面朝上的概率为 3/4 的硬币。从两枚硬币中挑出一枚，将这枚硬币掷 n 次。抛多少次足以让我们有 95% 的把握选择了哪种硬币？请写出计算过程。

Solution:

Solution. To guess which coin was picked, set a threshold t between $1/2$ and $3/4$. If the proportion of heads is less than the threshold, guess it was the fair coin; otherwise, guess the biased coin. Let the random variable H_n be the number of heads in the first n flips. We need to flip the coin enough times so that $\Pr(H_n/n > t) \leq 0.05$ if the fair coin was picked, and $\Pr(H_n/n < t) \leq 0.05$ if the biased coin was picked. A natural threshold to choose is $5/8$, exactly in the middle of $1/2$ and $3/4$.

H_n is the sum of independent Bernoulli variables, which each have variance $1/4$ for the fair coin and $3/16$ for the biased coin. Using Chebyshev's Inequality for the fair coin,

$$\begin{aligned}\Pr\left(\frac{H_n}{n} > \frac{5}{8}\right) &= \Pr\left(\frac{H_n}{n} - \frac{1}{2} > \frac{5}{8} - \frac{1}{2}\right) = \Pr\left(H_n - \frac{n}{2} > \frac{n}{8}\right)\\ &= \Pr\left(H_n - \mathrm{E}\left[H_n\right] > \frac{n}{8}\right) \leq \Pr\left(\left|H_n - \mathrm{E}\left[H_n\right]\right| > \frac{n}{8}\right)\\ &\leq \frac{\mathrm{Var}\left[H_n\right]}{(n/8)^2} = \frac{n/4}{n^2/64} = \frac{16}{n}\end{aligned}$$

For the biased coin, we have

$$\begin{aligned}\Pr\left(\frac{H_n}{n} < \frac{5}{8}\right) &= \Pr\left(\frac{3}{4} - \frac{H_n}{n} > \frac{3}{4} - \frac{5}{8}\right) = \Pr\left(\frac{3n}{4} - H_n > \frac{n}{8}\right)\\ &= \Pr\left(\mathrm{E}\left[H_n\right] - H_n > \frac{n}{8}\right) \leq \Pr\left(\left|H_n - \mathrm{E}\left[H_n\right]\right| > \frac{n}{8}\right)\\ &\leq \frac{\mathrm{Var}\left[H_n\right]}{(n/8)^2} = \frac{3n/16}{n^2/64} = \frac{12}{n}\end{aligned}$$

We are 95% confident if these are at most 0.05, which is satisfied if $n \geq 320$.

Because the variance of the biased coin is less that of the fair coin, we can do slightly better if make our threshold a bit bigger, to about 0.634, which gives 95% confidence with 279 coin flips.

图 1: Solution of Problem 3.

Problem 4. 对于一条马尔科夫链，定义其状态转移矩阵为 $\mathbf{P} = (p_{ij})_{i,j\in S}$，其中 S 为状态空间，

$$p_{ij} = P(X_{n+1} = j | X_n = i),$$

一个常返状态 i 的首次回归时间为

$$T_i = \min\{n > 0 : X_n = i | X_0 = i\}.$$

我们定义：

- 当 $\mathbb{E}[T_i] < \infty$，则称状态 i 是正常返 (positive recurrent) 的；
- 当 $P(T_i < \infty) = 1$ 但 $\mathbb{E}[T_i] = \infty$，则称状态 i 是零常返 (null recurrent) 的；
- 当 $P(T_i < \infty) < 1$，则称状态 i 是滑过 (transient) 的。

记 π_i 表示从状态 i 出发再回到 i 的平均回转时间，

(a) (10 分) 证明：$\pi_i = 1/\mathbb{E}(T_i)$；

(b) (10 分) 证明：不可约、有限状态的马尔科夫链中的所有常返状态都是正常返的；

(c) (20 分) 证明：不可约、正常返、非周期的马尔科夫链满足 $\pi = \pi P$，即马尔科夫链的平稳分布等于其极限分布。

Solution:

(a) 解法如图2。

(b) 解法如图3。

(c) 解法如图4。

Problem 5. (10 分) 股票市场预测与隐马尔可夫模型

在股票市场中，时间序列数据是非常重要的，因为股票价格的波动与时间密切相关。隐马尔可夫模型（hidden markov model, HMM) 是一种常用的建模时间序列数据的方法，它可以通过观测到的数据来推断预测出数据未来状态。

实验目的

本实验的目的是通过构建隐马尔可夫模型来预测未来的股票数据。具体来说，我们将使用过去的股票数据来训练隐马尔可夫模型，并使用该模型来预测未来的股票交易量。

实验要求

请基于该数据集，构建隐马尔可夫模型并预测 2018/12/29 - 2019/1/4（假设无休市）的股票交易量 (Volume)。该实验需提交实验报告与完整代码。

实验数据

本实验题的数据集包含了某支股票 730 天的价格数据，其中数据都是用日期格式来组织的。在这个数据集中，我们可以看到该股票数据随时间的波动情况。

数据字段：

Why $\pi_i = 1/\mathbb{E}(T_i)$?

Consider a Markov chain started from state j. Let S_k be the time till the k-th visit to state i. Then

$$S_k = T_{ji} + T_{ii}(1) + \ldots + T_{ii}(k-1)$$

Here

- T_{ji} = the first time the process visits state i from state j, and
- $T_{ii}(m)$ = the time between the mth and $(m+1)$st visit to state i.

Observe that $T_{ii}(1)$, $T_{ii}(2), \ldots T_{ii}(k-1)$ are i.i.d. and have the same distribution as T_i.

For k large, the Law of Large Numbers tells us

$$\frac{1}{k}[T_{ji} + T_{ii}(1) + T_{ii}(2) + \cdots + T_{ii}(k-1)] \approx \mathbb{E}(T_i)$$

i.e., the chain visits state i about k times in $k\mathbb{E}(T_i)$ steps.

We have just seen that in n steps, we expect about $n\pi(i)$ visits to the state i. Hence setting $n = k\mathbb{E}(T_i)$, we get the relation

$$\pi_i = 1/\mathbb{E}(T_i).$$

图 2: Solution of Exercise 4.a

In a finite-state Markov chain all recurrent states are positive recurrent.

Proof.

It suffices to consider irreducible Markov chains only since a Markov chain restricted to one of its recurrent class is also a Markov chain.

Recall an irreducible Markov chain must be recurrent. Also recall that positive/null recurrence is a class property. Thus if one state is null recurrent, then all states are null recurrent. However, since $\sum_{j\in\mathfrak{X}} P_{ij}^{(n)} = 1$. As there are only finite number of states, it is impossible that $\lim_{n\to\infty} P_{ij}^{(n)} = 0$ for all $j \in \mathfrak{X}$. Thus no state can be null recurrent.

图 3: Solution of Exercise 4.b

Theorem 2.4 (**Fundamental theorem for Markov Chains**) *An irreducible, ergodic Markov chain has a unique stationary distribution π. The limiting distribution exists and is equal to π.*

Proof Since the chain is ergodic, it is non-null recurrent which implies from above that $\pi_j = \lim_{n\to\infty} p_{ij}(n) > 0 \,\forall i$ and $\sum \pi_j = 1$. Now we consident for any M $\sum_{i=0}^{M} p_{ij}(n) \leq \sum_{i=0}^{\infty} p_{ij}(n) = 1$. Now, letting $n \to \infty$ we get that $\sum_{i=0}^{M} \pi_i \leq 1 \,\forall M$ which implies from above that the same is true for the inifinte case $\sum_{i=0}^{\infty} \pi_i \leq 1$. Now we consider the probability of moving from i to j in $n+1$ steps so $p_{ij}(n+1) = \sum_{k=0}^{\infty} p_{ik}(n) p_{kj} \geq \sum_{k=0}^{M} p_{ik}(n) p_{kj} \,\forall M$. Now we again can let $n \to \infty$ which implies that $\pi_i \geq \sum_{k=0}^{M} \pi_k p_{kj} \,\forall M$ which implies that $\pi_i \geq \sum_{k=0}^{\infty} \pi_k p_{kj}$. Now we assume the inequality is strict for some i which leads to the following contradiction:

$$\begin{aligned} \sum_{i=0}^{\infty} \pi_i &\geq \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \pi_k p_{ki} \\ &= \sum_{k=0}^{\infty} \pi_k \sum_{i=0}^{\infty} p_{ki} \\ &= \sum_{k=0}^{\infty} \pi_k . \end{aligned}$$

Thus we come to the conclusion that $\pi_i = \sum_{k=0}^{\infty} \pi_k p_{ki} \,\forall i$. Now we can consider $\tilde{\pi}_i = \pi_i / \sum_{k=0}^{\infty} \pi_k$ to be a stationary distribution. So we have shown existence. Now to show uniqueness we consider the following:

$$\tilde{\pi}_i = \mathbb{P}(X_n = i) = \sum_{j=0}^{\infty} \mathbb{P}(X_n = i | X_0 = j) \mathbb{P}(X_0 = j) = \sum_{j=0}^{\infty} p_{ji}(n) \tilde{\pi}_j .$$

So we have that $\tilde{\pi}_i \geq \sum_{j=0}^{M} p_{ji}(n) \tilde{\pi}_j$and taking $M, n \to \infty$ we get that $\tilde{\pi}_i \geq \sum_{j=0}^{\infty} \tilde{\pi}_j \pi_i = \pi_i$ but we know that from the transition matrix that $p_{ji}(n) \leq 1$ and so $\tilde{\pi}_i \leq \sum_{j=0}^{M} p_{ji}(n) \tilde{\pi}_j + \sum_{j=M+1}^{\infty} \tilde{\pi}_j \,\forall M$ and so taking $n \to \infty$we get $\tilde{\pi}_i \leq \sum_{j=0}^{M} \pi_i \tilde{\pi}_j + \sum_{j=M+1}^{\infty} \tilde{\pi}_j \,\forall M$. Now we know that $\tilde{\pi}$is a stationary distribution so it sums up to 1, and so we let $M \to \infty$ and we get $\tilde{\pi}_i \leq \sum_{j=0}^{\infty} \pi_i \tilde{\pi}_j = \pi_i$ which implies that the stationary distribution is unique.

The above process thus shows the existence of a limiting distribution and so we now know that an ergodic chain converges to its stationary distribution.

图 4: Solution of Exercise 4.c

- data: 开盘时间
- open: 开盘价
- high: 当日最高价
- low: 当日最低价
- close: 收盘价
- volume: 当日交易量

数据链接：https://rec.ustc.edu.cn/share/57b34db0-c721-11ed-b9b0-99e32b708b65
获取密码：xwvz

Solution:

```
#隐马尔可夫预测
import datetime
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.finance import quotes_historical_yahoo_ochl
from hmmlearn.hmm import GaussianHMM

#从雅虎财经获取股票报价,在matplotlib库中有一个函数可以直接加载:
quotes=quotes_historical_yahoo_ochl("INTC",datetime.date(1994,4,5),
         datetime.date(2015,7,3))
# 提取需要的数值    每个报价包含6个值。下面提取相关数据，如股票的收盘价和一定时期内股票的成交量
dates=np.array([quote[0] for quote in quotes],dtype=np.int)
closing_values=np.array([quote[2] for quote in quotes])
volume_of_shares=np.array([quote[5] for quote in quotes])[1:]

# 计算每天收盘价的变化率
diff_percentage=100.0*np.diff(closing_values)/closing_values[:-1]
dates=dates[1:]

# 将变化率与交易量组合起来
X=np.column_stack([diff_percentage,volume_of_shares])
# 创建并训练高斯HMM模型
print "\nTraining HMM.... "
model=GaussianHMM(n_components=5,covariance_type="diag" ,n_iter=1000)
model.fit(X)
```

图 5: Code for solution

COMP6003P.02 计算机应用数学 Spring 2023 Homework 2

教师：侯嘉慧 due: May. 21, 23:59

Problem 1. （10 分）给定贝叶斯网络如图1所示，完成以下问题。

(a) 绘制该网络对应的因子图，并给出联合分布概率公式（形如 $g(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)$ $f_3(x_1, x_2, x_3)$）；

(b) 给出 Cancer 对应的边缘概率分布公式；

(c) 以 Cancer 对应的边缘概率分布为例，对比直接计算和使用 Belief Propagation 算法（或称作 Sum-Product Message Passing）的计算代价之间的差异（即比较乘法加法次数）。

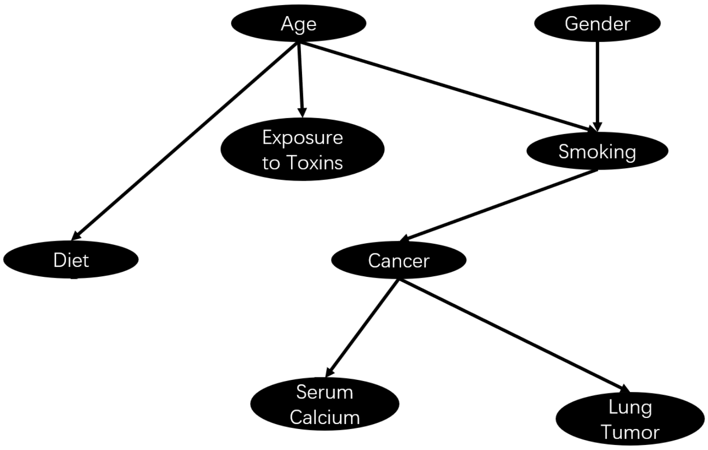
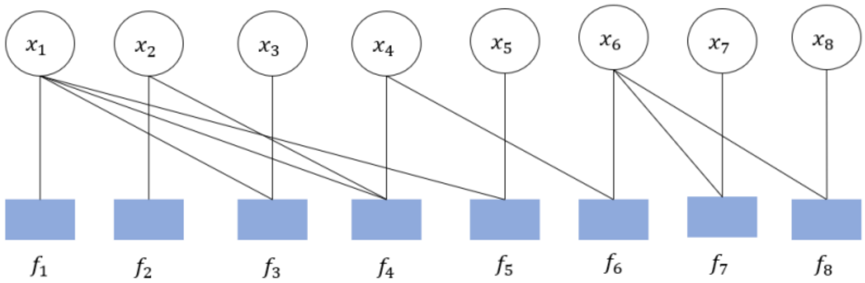


图 1: 贝叶斯网络

Solution:

令 Age: x_1, Gender: x_2, Exposure to Toxins: x_3, Smoking: x_4, Diet: x_5, Cancer: x_6, Serum Calcium: x_7, Lung Tumor: x_8

a) 绘制出的因子图如下所示：



联合概率分布如下:

$$g(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8)$$
$$= f_1(x_1)f_2(x_2)f_3(x_1,x_3)f_4(x_1,x_2,x_4)f_5(x_1,x_5)f_6(x_4,x_6)f_7(x_6,x_7)f_8(x_6,x_8)$$

b) Cancer：x_6对应的边缘概率公式如下所示

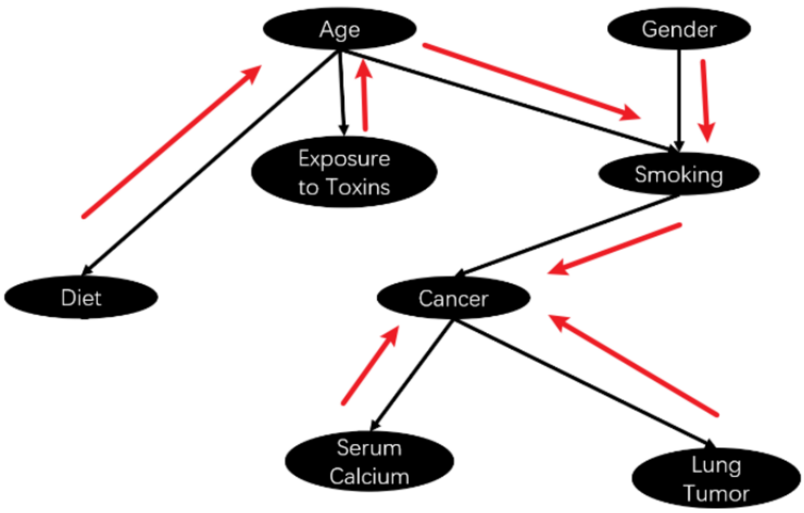
$$P(x_6) = \sum_{x_8}\sum_{x_7}\sum_{x_5}\sum_{x_4}\sum_{x_3}\sum_{x_2}$$

$$\sum_{x_1} f_1(x_1)f_2(x_2)f_3(x_1,x_3)f_4(x_1,x_2,x_4)f_5(x_1,x_5)f_6(x_4,x_6)f_7(x_6,x_7)f_8(x_6,x_8)$$

c) 考虑用信念传播算法优化，则有

变量的边缘概率等于所有与他相连的函数传递过来的消息的积。

$$f(x_6) = m_{\rightarrow}x_6 * m_{\leftarrow}x_6$$



所以

$$m_{\leftarrow}x_1 = \sum_{x_3,x_5} f_3(x_1,x_3)\cdot f_5(x_1,x_5)\cdot f_1(x_1)$$

$$m_{\rightarrow}x_2 = f_2(x_2)$$

$$m_{\leftarrow}x_4 = \sum_{x_1,x_2} f_4(x_1,x_2,x_4)\cdot m_{\rightarrow}x_2\cdot m_{\leftarrow}x_1$$

$$= \sum_{x_1,x_2} f_4(x_1,x_2,x_4)\cdot f_2(x_2)\cdot\left(\sum_{x_3,x_5} f_3(x_1,x_3)\cdot f_5(x_1,x_5)\cdot f_1(x_1)\right)$$

$$m_{\leftarrow}x_6 = \sum_{x_4} f_6(x_4,x_6)\cdot m_{\leftarrow}x_4$$

$$= \sum_{x_4} f_6(x_4,x_6)\cdot\left[\sum_{x_1,x_2} f_4(x_1,x_2,x_4)\cdot f_2(x_2)\cdot\left(\sum_{x_3,x_5} f_3(x_1,x_3)f_5(x_1,x_5)\cdot f_1(x_1)\right)\right]$$

$$m_{\rightarrow}x_6 = \sum_{x_7,x_8} f_7(x_6,x_7)\cdot f_8(x_6,x_8)$$

所以, $\bar{f}_6(x_6) = m_{\rightarrow}x_6 * m_{\leftarrow}x_6$

$$= \left(\sum_{x_7,x_8} f_7(x_6,x_7)\cdot f_8(x_6,x_8)\right)\cdot$$

$$\left(\sum_{x_4} f_6(x_4,x_6)\cdot\left[\sum_{x_1,x_2} f_4(x_1,x_2,x_4)\cdot f_2(x_2)\cdot\left(\sum_{x_3,x_5} f_3(x_1,x_3)f_5(x_1,x_5)\cdot f_1(x_1)\right)\right]\right)$$

总的乘法次数： 左 + 右 + 1

左： $k(1+m_4)$

m_4： $k^2(2+m_2+m_1)$

m_2： 0次

m_1 ： $2k^2$

∴ 左 ： $k[1+k^2(2+2k^2)]$

$= k+2k^3+2k^5$

右 ： k^2

∴ 总的乘法次数 $2k^5+2k^3+k^2+k+1$

总的加法次数 $7(k-1)$

Problem 2. （10 分）请证明：高维空间随机取两个向量，它们几乎正交。

Hint 1. 原问题可以转换为证明不等式：

$$Pr\left[|cos(\theta_{x,y})| > \sqrt{\frac{logc}{n}}\right] < \frac{1}{c},$$

其中，x, y 是从 $\{-1,1\}^n$ 中随机选取的两个向量，它们的夹角是 $\theta_{x,y}$。

Solution: 利用切诺夫界

$$Pr(|X| > t) < e^{-(\frac{t}{\sigma})^2}$$

Problem 3. （20 分）给定图 G（如图2，图中数值为边权值），图切割将其分割成多个互不连通的子图。请使用谱聚类算法将图 G 聚类成 $k = 3$ 类，使得：

(a) RatioCut 最小；

(b) NormalizedCut 最小。

并给出计算过程，以及对应的 RatioCut 值和 NormalizedCut 值。

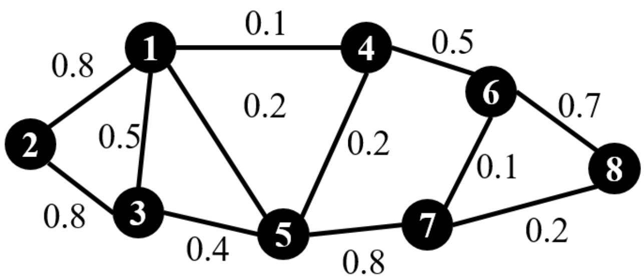


图 2: 图 G

Solution:

(a) RatioCut 值的计算方式：

$$\text{RatioCut}(A_1, A_2, ..., A_k) = \frac{1}{2}\sum_{i=1}^{k}\frac{W(A_i, \bar{A}_i)}{|A_i|},$$

其中 $\bar{A}_i$ 为 A_i 的补集，$W(A_i,\bar{A}_i)=\sum_{i\in A_i,j\in\bar{A}_i} w_{ij}$, $|A_i|$ 为 A_i 的节点个数。

求解算法：

Step 1. 求解 Laplacian 矩阵：$L=D-W$, 其中 $D=\mathrm{diag}d_1,d_2,...d_n$ 为所有节点度的对角矩阵。

Step 2. 对 L 进行特征值分解，并取 k 个（k 为需要分类的类别数目，在此 $k=3$）最小特征设置对应的特征向量。

Step 3. 将求解的 K 个特征向量归一化后构成新的矩阵, 对该矩阵进行 K-means 聚类处理。

聚类结果：$\{1,2,3\},\{4,6,8\},\{5,7\}$

RatioCut ≈ 0.4917。

(b) NormalizedCut 值的计算方式：

$$\mathrm{NCut}(A_1,A_2,...,A_k)=\frac{1}{2}\sum_{i=1}^{k}\frac{W(A_i,\bar{A}_i)}{vol(A_i)},$$

其中 $\bar{A}_i$ 为 A_i 的补集，$vol(A_i)=\sum_{i\in A_i} d_i$。

求解算法：

Step 1. 求解 Laplacian 矩阵：$L=D^{-\frac{1}{2}}(D-W)D^{-\frac{1}{2}}$, 其中 $D=\mathrm{diag}\{d_1,d_2,...d_n\}$ 为所有节点度的对角矩阵。

Step 2. 对 L 进行特征值分解，并取 k 个（k 为需要分类的类别数目，在此 $k=3$）最小特征设置对应的特征向量。

Step 3. 将求解的 K 个特征向量归一化后构成新的矩阵, 对该矩阵进行 K-means 聚类处理。

聚类结果：$\{1,2,3\},\{4,6,8\},\{5,7\}$

NCut ≈ 0.3751。

Problem 4. （10 分）参数未知的正态分布生成了如下 5 组数据：-12.2119174, 7.62328757, 9.60884573, -8.36968016, 2.42378373。请使用极大似然估计方法估计出正态分布的参数，写出计算过程。

Solution: 答案如图3所示。

$$L(\mu,\sigma|x_1,\cdots,x_n) = L(\mu,\sigma|x_1) * L(\mu,\sigma|x_2) \cdots L(\mu,\sigma|x_n)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}}$$

$$\ln L(\mu,\sigma|x_1,\cdots,x_n) = n \ln\left[(2\pi\sigma^2)^{-\frac{1}{2}}\right] - \frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}$$

$$= -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \ln L(\mu,\sigma|x_1,\cdots,x_n)}{\partial\mu} = \frac{1}{\sigma^2}\left[x_1+\cdots+x_n - n\mu\right]$$

令其$=0$ 得 $\mu = \frac{x_1+\cdots+x_n}{n}$

$$\frac{\partial \ln L(\mu,\sigma|x_1,\cdots,x_n)}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\left[(x_1-\mu)^2+\cdots+(x_n-\mu)^2\right]$$

令其$=0$ 得 $\sigma = \sqrt{\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{n}}$

图 3: 第四题答案

Problem 5. （20 分）HITS 全称是 Hyperlink-Induced Topic Search，即超链接诱导主题搜索，是一个链接分析方法，由康奈尔大学的 Kleinberg 教授提出。HITS 算法中有 2 个比较重要的概念，一个是“Authority”页面，另一个是“Hub”页面。“Authority”，即权威，所有权威页面，是指页面本身的质量比较高，比如，百度搜索或者谷歌搜索首页。“Hub”，即枢纽，表示本页面指向的其他很多高质量的“Authority”页面。一个好的“Authority”会被很多好的“Hub”页面指向。一个好的“Hub”页面会指向很多好的“Authority”页面。权威值和枢纽值是互相依存、互相影响的。网页的权威值是所有指向它的网页的枢纽值之和：

$$auth(p) = \Sigma_{q \in P_{to}} hub(q),$$

网页的枢纽值是所有它指向网页的权威值之和：

$$hub(p) = \Sigma_{q \in P_{from}} auth(q).$$

HITS 算法流程如下：

1. 初始化：将各节点的权威值和枢纽值均设为 1。
2. 更新节点的权威值;
3. 更新节点的枢纽值;
4. 将权威值和枢纽值归一化，即权威值和枢纽值分别和为 1;
5. 重复 2-4 步骤，直至最终收敛。

参考：HITS 论文 。

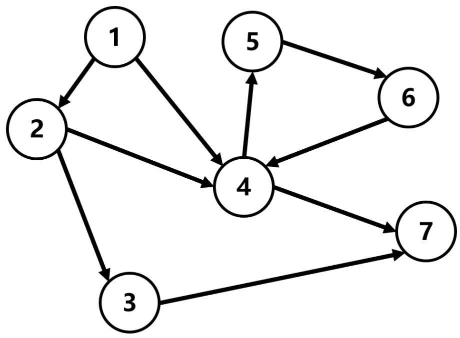


图 4: 网页关系图

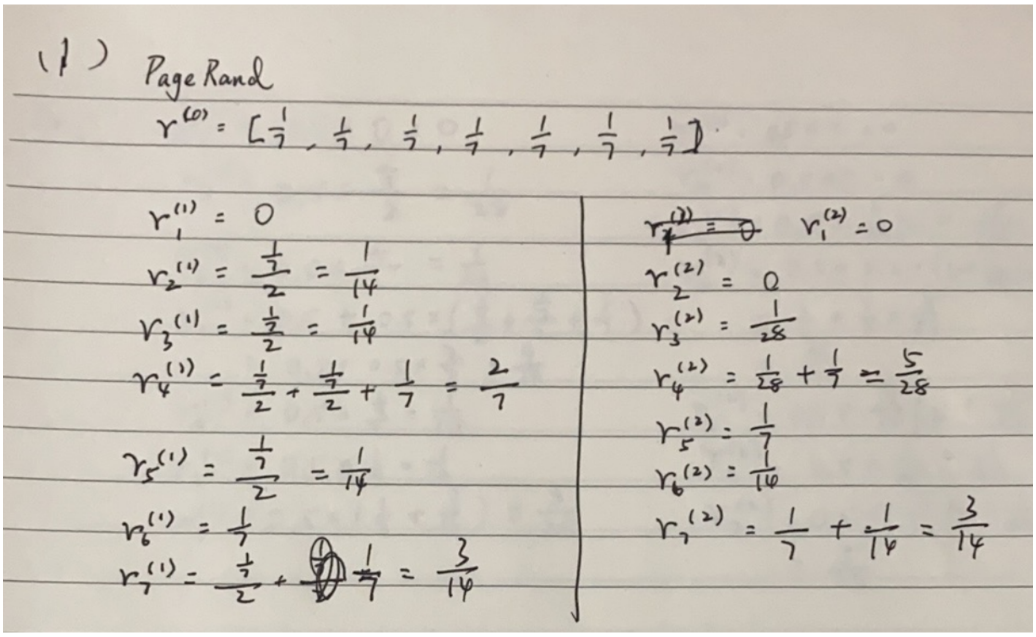
(a) 根据如图 4 所示的网络关系，每个节点的 PageRank 初始值都是 $\frac{1}{7}$，写出 PageRank 算法前两轮迭代的过程。

(b) 在第 (a) 小题的基础上，假设停止游走的概率 $p = 0.5$，请写出对于节点 4 的 Personalized PageRank 算法前两轮迭代的过程。

(c) 根据如图 4 所示的网络关系，每个节点的初始权威值和枢纽值均设为 1，写出 HITS 算法前两轮迭代的过程。

(d) 比较 PageRank、Personalized PageRank 和 HITS 这三个算法的作用和优缺点。

Solution: 答案如图5、图6、图7所示



(1) PageRand

$r^{(0)} = [\frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}]$

$r_1^{(1)} = 0$

$r_2^{(1)} = \frac{\frac{1}{7}}{2} = \frac{1}{14}$

$r_3^{(1)} = \frac{\frac{1}{7}}{2} = \frac{1}{14}$

$r_4^{(1)} = \frac{\frac{1}{7}}{2} + \frac{\frac{1}{7}}{2} + \frac{1}{7} = \frac{2}{7}$

$r_5^{(1)} = \frac{\frac{1}{7}}{2} = \frac{1}{14}$

$r_6^{(1)} = \frac{1}{7}$

$r_7^{(1)} = \frac{\frac{1}{7}}{2} + \frac{1}{7} = \frac{3}{14}$

$r_1^{(2)} = 0$

$r_2^{(2)} = 0$

$r_3^{(2)} = \frac{1}{28}$

$r_4^{(2)} = \frac{1}{28} + \frac{1}{7} = \frac{5}{28}$

$r_5^{(2)} = \frac{1}{7}$

$r_6^{(2)} = \frac{1}{14}$

$r_7^{(2)} = \frac{1}{7} + \frac{1}{14} = \frac{3}{14}$

图 5: PageRank 前两轮的迭代

Problem 6. （20 分）证明图的拉普拉斯矩阵存在以下性质：

(a) 假设 G 是 d-regular graph，证明 0 是拉普拉斯矩阵的特征值；

(b) 假设 G 是 d-regular graph，证明当且仅当 G 是连通时，其拉普拉斯矩阵的特征值 0 的重数恰好为 1；

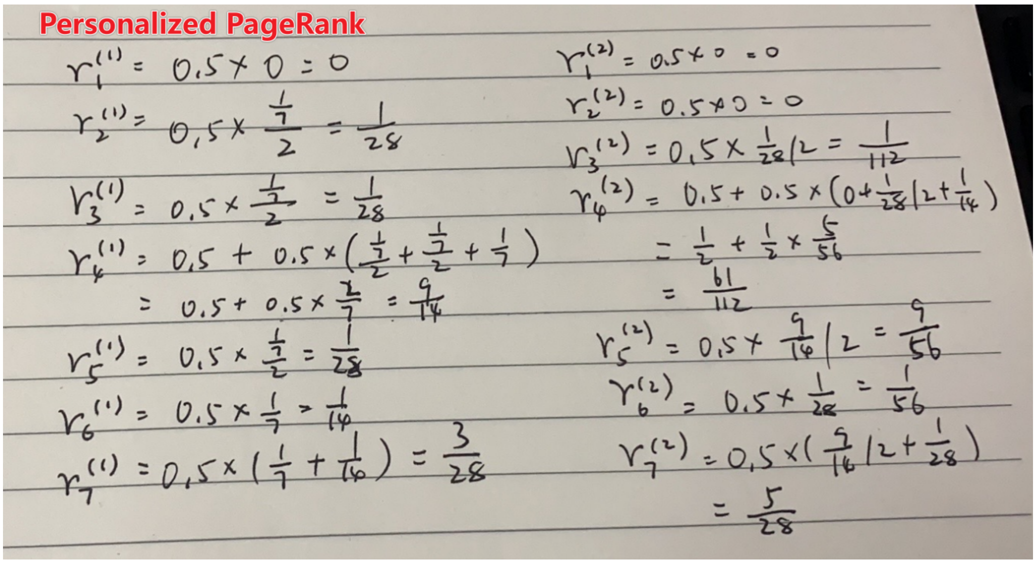


图 6: Personalized PageRank 前两轮的迭代

(c) 假设分块对角矩阵 $B=\begin{bmatrix} A_1 & 0 & ... & 0 \\ 0 & A_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & A_n \end{bmatrix}$，证明 $detB = detA_1 detA_2 ... detA_n$；

(d) 假设 G 是一个有非负权重的无向图，证明其拉普拉斯矩阵 L 的特征值 0 的重数 k 等于图的联通分量的个数。

Hint 2. 基于 *(a)-(c)* 的结论证明 *(d)*

Solution:

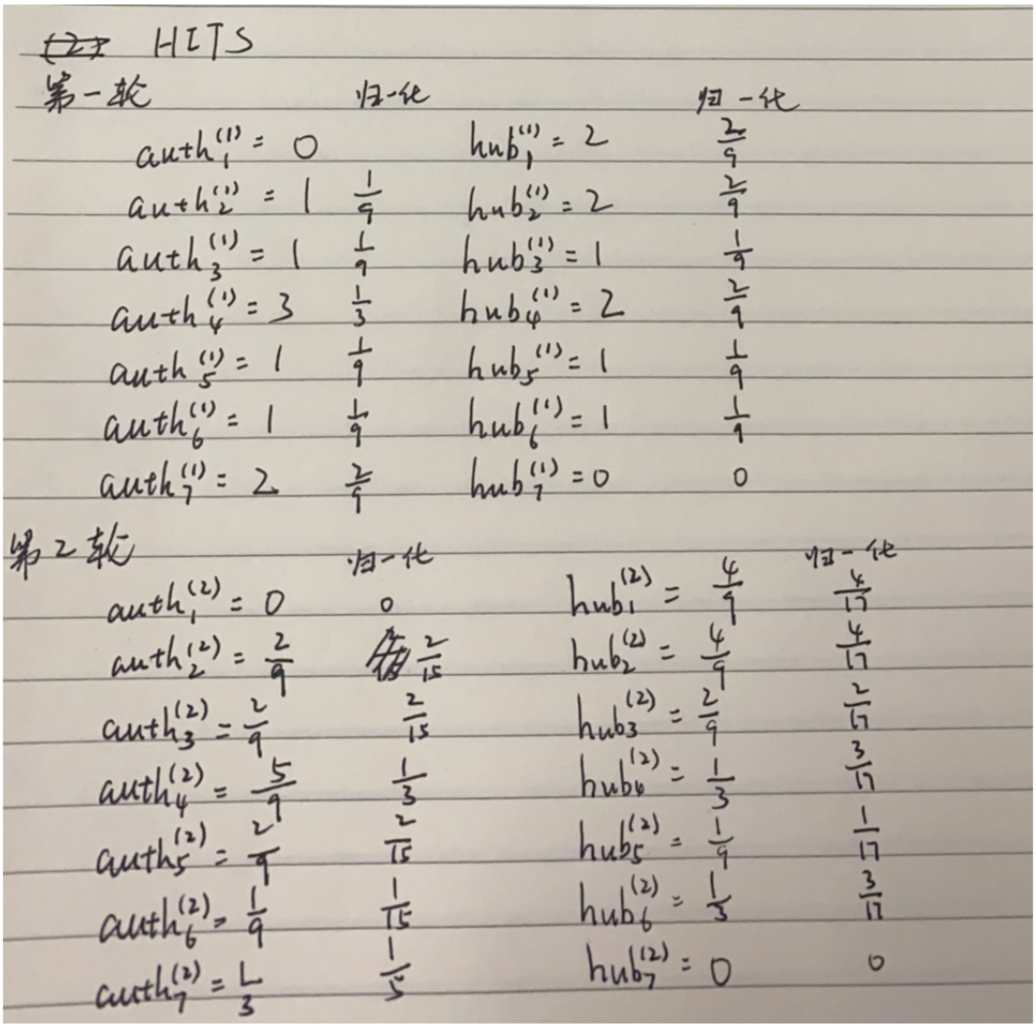
(a), (b), (d) 答案如图8, **??**, **??**.

(c) 由行列式的计算定义可得。

Problem 7. （10 分）旅行者在如图 11 所示的路线中旅行，10 号节点是目的地。旅行者可能出现在各个节点上，请使用强化学习为他提供前进的路线。

实验目的

本实验的目的是通过构建强化学习模型进行自动寻路。具体来说，根据给定的路线图和当前位置，生成下一步移动的节点，直到行至终点。



~~(2)~~ HITS

第一轮

	归一化		归一化
$auth_1^{(1)} = 0$		$hub_1^{(1)} = 2$	$\frac{2}{9}$
$auth_2^{(1)} = 1$	$\frac{1}{9}$	$hub_2^{(1)} = 2$	$\frac{2}{9}$
$auth_3^{(1)} = 1$	$\frac{1}{9}$	$hub_3^{(1)} = 1$	$\frac{1}{9}$
$auth_4^{(1)} = 3$	$\frac{1}{3}$	$hub_4^{(1)} = 2$	$\frac{2}{9}$
$auth_5^{(1)} = 1$	$\frac{1}{9}$	$hub_5^{(1)} = 1$	$\frac{1}{9}$
$auth_6^{(1)} = 1$	$\frac{1}{9}$	$hub_6^{(1)} = 1$	$\frac{1}{9}$
$auth_7^{(1)} = 2$	$\frac{2}{9}$	$hub_7^{(1)} = 0$	0

第2轮

	归一化		归一化
$auth_1^{(2)} = 0$	0	$hub_1^{(2)} = \frac{4}{9}$	$\frac{4}{17}$
$auth_2^{(2)} = \frac{2}{9}$	$\frac{2}{15}$	$hub_2^{(2)} = \frac{4}{9}$	$\frac{4}{17}$
$auth_3^{(2)} = \frac{2}{9}$	$\frac{2}{15}$	$hub_3^{(2)} = \frac{2}{9}$	$\frac{2}{17}$
$auth_4^{(2)} = \frac{5}{9}$	$\frac{1}{3}$	$hub_4^{(2)} = \frac{1}{3}$	$\frac{3}{17}$
$auth_5^{(2)} = \frac{2}{9}$	$\frac{2}{15}$	$hub_5^{(2)} = \frac{1}{9}$	$\frac{1}{17}$
$auth_6^{(2)} = \frac{1}{9}$	$\frac{1}{15}$	$hub_6^{(2)} = \frac{1}{3}$	$\frac{3}{17}$
$auth_7^{(2)} = \frac{1}{3}$	$\frac{1}{5}$	$hub_7^{(2)} = 0$	0

图 7: HITS 前两轮的迭代

Lemma 3.5. *Let G be a d-regular graph. Then 0 is an eigenvalue for the Laplacian matrix of G.*

Proof. If G is a d-regular graph the sum of the entries in row i gives the degree of vertex i so we have,

$$A\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}=\begin{pmatrix}d\\d\\\vdots\\d\end{pmatrix}$$

Then one can see that

$$(D-A)\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}=\begin{pmatrix}d-d\\d-d\\\vdots\\d-d\end{pmatrix}=\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}=0\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}$$

图 8: 第六题 (a) 答案

实验要求

请基于给出的路线图，生成移动策略。使用课程中强化学习模型的任意一种即可，编程实现。该实验需提交实验报告与完整代码。

生成策略的一个例子：在节点 4，强化学习模型选择前往节点 5；在节点 5，强化学习模型选择前往节点 8；在节点 8，强化学习模型选择前往节点 10，达到终点。

Lemma 3.6. *Let $x \in \mathbb{R}^n$ where $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Let L denote the Laplacian matrix of graph G and E denote the set of edges in G. Then $x^T Lx = \sum\limits_{[i,j]\in E} (x_i - x_j)^2$*

Proof. By definition of the Laplacian we have,

$$x^T Lx = x^T(D - A)x = x^T Dx - x^T Ax$$

Expanding this out we get

$$\begin{pmatrix} x_1 \dots x_n \end{pmatrix} \begin{bmatrix} deg(1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & deg(n) \end{bmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \begin{pmatrix} x_1 \dots x_n \end{pmatrix} (a_{ij}) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} =$$

$$\sum_i deg(i)x_i^2 - \begin{pmatrix} x_1 \dots x_n \end{pmatrix} (a_{ij}) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Recall that $a_{ij} = 1$ if there is an edge between vertex i and j and $a_{ij} = 0$ otherwise, so $\begin{pmatrix} x_1 \dots x_n \end{pmatrix} (a_{ij}) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum\limits_{[i,j]\in E} a_{ij} x_i x_j = \sum\limits_{[i,j]\in E} 2xixj$. Thus $x^T Lx = \sum\limits_i deg(i)x_i^2 - \sum\limits_{[i,j]\in E} a_{ij} xixj = \sum\limits_{[i,j]\in E} (x_i^2 - 2xixj + x_j^2) = \sum\limits_{[i,j]\in E} (x_i - x_j)^2$ $\square$

图 9: 第六题 (b) 答案

Theorem 3.9. *Let G be a d-regular graph. The algebraic multiplicity of eigenvalue 0 for the Laplacian matrix is exactly 1 iff G is connected.*

Proof. We prove the first direction by contrapositive. Suppose G is not connected. Then it is possible to write G's adjacency matrix so that it is block diagonal. To do this, label vertices contained in the same connected component of G consecutively so that vertices $\{1, 2, \ldots, a-1, a\}$ are part of one connected component; vertices $\{a+1, a+2, \ldots, a+b-1, a+b\}$ are part of another connected component and so on. Let A_1 denote the adjacency matrix for the connected component with vertices $\{1, 2, \ldots, a-1, a\}$, A_2 denotes the adjacency matrix for the connected component with vertices $\{a+1, a+2, \ldots, a+b-1, a+b\}$, and so on. Then our adjacency matrix for G can be written as

$$\begin{bmatrix} A_1 & \ldots & & & \\ \vdots & A_2 & \ldots & & \\ & \vdots & \ddots & & \\ & & & \ddots & \vdots \\ & & & \ldots & A_r \end{bmatrix}$$

Since there are no edges between these connected components, all values outside the A_is will be 0, so the matrix is block diagonal. Note that the Laplacian is also block diagonal. Let the blocks be denoted $L_1, \ldots, L_r$. By Lemma 3.5 each block matrix in the Laplacian has eigenvalue 0. That is, for any i, $det(xI - L_i)$ has a $(x-0)$ term. By Lemma 3.7 $det(xI - L) = det(xI - L_1) \ldots det(xI - L_r)$, so $det(xI - L)$ has more than one $(x-0)$ term so the algebraic multiplicity is greater than 1. Now we prove the second direction by contradiction. Suppose G is connected but the algebraic multiplicity of eigenvalue 0 is greater than 1. Then by Corollary 2.13, there exist at least two linearly independent eigenvectors v_1, v_2 with eigenvalue 0. Then $v_1^T L v_1 = 0$ and $v_2^T L v_2 = 0$ so by Lemma 3.6,

$$\sum_{[i,j] \in E} (v_{i1} - v_{j1})^2 = 0$$

and

$$\sum_{[i,j] \in E} (v_{i2} - v_{j2})^2 = 0$$

Since all entries in the sum are positive, this can only be zero when $\forall [i,j] \in E$, $v_{i1} = v_{j1}$ and $v_{i2} = v_{j2}$. This also implies that if there is a path between two vertices m and n, then $v_{m1} = v_{n1}$ and $v_{m2} = v_{n2}$ even if $[m,n] \notin E$. But if G is connected there is a path between any pair of vertices. Therefore v_1 and v_2 are constant vectors (every entry is the same). Thus v_1 and v_2 are scalar multiples of each other. But v_1 and v_2 are linearly independent which is a contradiction. □

图 10: 第六题 (d) 答案

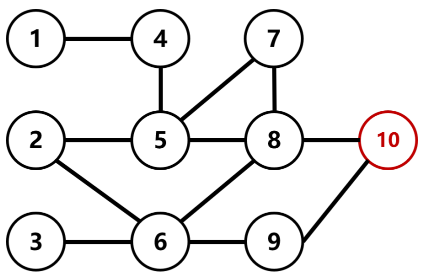


图 11: 路线图

Problem 1. （20 分）计算矩阵 $B = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$ 的奇异值分解 $B = U\Sigma V^T$。给出计算过程。

Solution:

$$U = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \end{pmatrix}$$

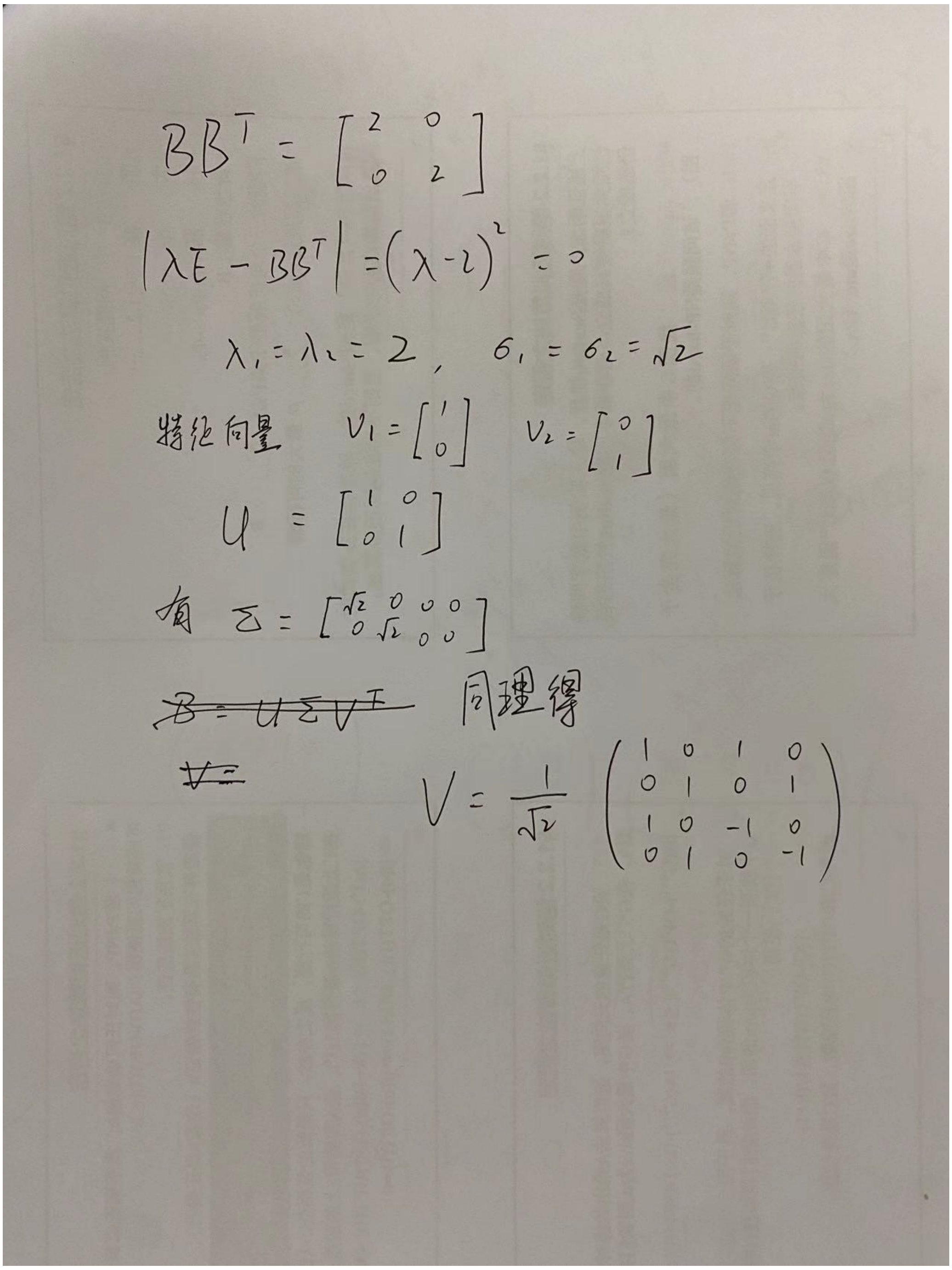
$$V = \begin{pmatrix} -1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} / \sqrt{2}$$

详细见图1。

Problem 2. （20 分）回答以下问题：

(a) 存在 10 个点 $(0,1),(1,11),(2,31),(3,60),(4,101),(5,146),(6,207),(7,279),(8,356),(9,445)$。使用形如 $\theta_0 + \theta_1 x^1 + \theta_2 x^2$ 的多项式拟合以上数据。

(b) 已知正例点 $x_1 = (1,2)^T, x_2 = (2,3)^T, x_3 = (3,3)^T$，负例点 $x_4 = (2,1)^T, x_5 = (3,2)^T$，试求最大间隔分离超平面和分类决策函数，并在图上画出分离超平面、间隔边界及支持向量。



$$BB^T = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$|\lambda E - BB^T| = (\lambda - 2)^2 = 0$$

$\lambda_1 = \lambda_2 = 2$, $\sigma_1 = \sigma_2 = \sqrt{2}$

特征向量 $v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

有 $\Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \end{bmatrix}$

$\cancel{B = U\Sigma V^T}$ 同理得

$\cancel{V=}$

$$V = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$$

图 1: 第一题答案

Solution:

利用$f(x) = \theta_2 x^2 + \theta_1 x + \theta_0$拟合函数，令$L = \sum_0^9 (f(x_i) - y_i)^2$，令$L$最小，则有

$$\frac{\partial L}{\partial \theta_2} = 0$$

$$\frac{\partial L}{\partial \theta_1} = 0$$

$$\frac{\partial L}{\partial \theta_0} = 0$$

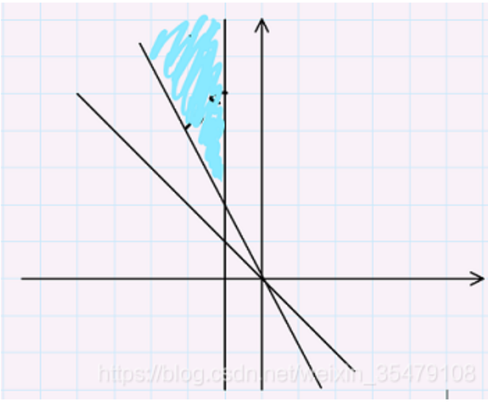
代入求得：

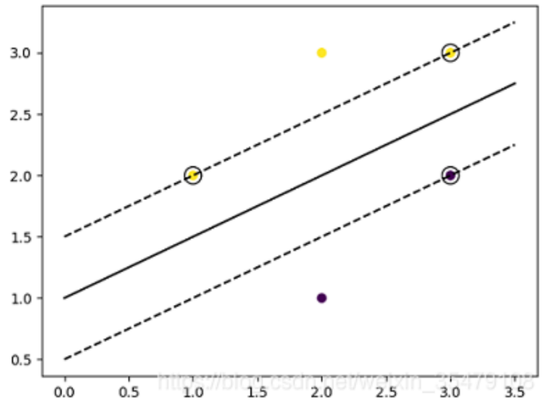
$$\begin{aligned}
\theta_2 &= \frac{(\overline{xy} - \bar{x} \cdot \bar{y})\left(\overline{x^3} - \bar{x} \cdot \overline{x^2}\right) - \left(\overline{x^2 y} - \overline{x^2} \cdot \bar{y}\right)\left(\overline{x^2} - (\bar{x})^2\right)}{\left(\overline{x^3} - \bar{x} \cdot \overline{x^2}\right)^2 - \left(\overline{x^4} - \left(\overline{x^2}\right)^2 \cdot \overline{x^2}\right)\left(\overline{x^2} - (\bar{x})^2\right)} \\
\theta_1 &= \frac{\overline{xy} - \bar{x} \cdot \bar{y} - \hat{a}\left(\overline{x^3} - \bar{x} \cdot \overline{x^2}\right)}{\overline{x^2} - (\bar{x})^2} \\
\theta_0 &= \bar{y} - \theta_2 \overline{x^2} - \theta_1 \bar{x}
\end{aligned}$$

将原始数据带入后得：$\theta_2 = 4.96590909, \theta_1 = 4.6219697, \theta_0 = 1.37272727$

手动计算：根据题意，得到目标函数即约束条件

$$
\begin{aligned}
&\min \frac{1}{2}\|w_1^2+w_2^2\| \\
s.t.\quad & w_1+2w_2+b \geq 1 \qquad (1)\\
& 2w_1+3w_2+b \geq 1 \qquad (2)\\
& 3w_1+3w_2+b \geq 1 \qquad (3)\\
& -2w_1-w_2-b \geq 1 \qquad (4)\\
& -3w_1-2w_2-b \geq 1 \qquad (5)
\end{aligned}
$$





以$w_1 . w_2$为坐标轴找到可行域，目标函数即求到原点距离最小的点，也就是$w=[-1,2]$,对于正例点$b \geq -2$，对于负例点$b \leq -2$，所以$b=-2$。

最大间隔分离超平面为 $-x_1+2x_2-2=0$

分类决策函数为 $sign(-x_1+2x_2-2)$

Problem 3. （20 分）现有如下数据（2 个维度），$X = \begin{bmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{bmatrix}$，其中，前三列和后两列分属两个类别。请分别使用 PCA、LDA 将数据降到 1 维，给出详细计算过程并分析两种方法的优劣。

Solution:

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** $C = \frac{1}{m} X X^T$
- **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**

LDA 算法

- **输入：数据集**$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$**，其中任意样本**$x_i$**为n维向量，** $y_i \in \{C_1, C_2, \dots, C_k\}$**。**
- **输出：降维到**d**维后的样本集D′**
- **1) 计算类内散度矩阵**S_w
- **2) 计算类间散度矩阵**S_b
- **3) 计算矩阵**$S_w^{-1} S_b$
- **4）计算**$S_w^{-1} S_b$**的最大的d个特征值和对应的d个特征向量**$(w_1, w_2, \dots w_d)$**，得到投影矩阵**W
- **5) 对样本集中的每一个样本特征**x_i**，转化为新的样本**$z_i = W^T x_i$
- **6) 得到输出样本集**$D = \{(z_1, y_1), (z_2, y_2), \dots, (z_m, y_m)\}$

利用 PCA 降维，如下：

数据X有 5 个样本，每个样本有两个特征。

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

由于数据的每一行都为 0，因此可以省略零均值化的步骤

则其协方差矩阵为：

$$C=\frac{1}{m}XX^T=\frac{1}{5}\begin{bmatrix}-1 & -1 & 0 & 2 & 0\\ -2 & 0 & 0 & 1 & 1\end{bmatrix}\begin{bmatrix}-1 & -2\\ -1 & 0\\ 0 & 0\\ 2 & 1\\ 0 & 1\end{bmatrix}=\begin{bmatrix}\frac{6}{5} & \frac{4}{5}\\ \frac{4}{5} & \frac{6}{5}\end{bmatrix}$$

令$|C-\lambda I|=0$，可以求得特征值和对应的特征向量为：

$$(\frac{6}{5}-\lambda)^2-\frac{16}{25}=0$$

$$\lambda_1=2,\lambda_2=\frac{2}{5}$$

$$c_1=[1,1]^T c_2=[-1,1]^T$$

将其标准化可得：

$$c_1=[\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}]^T c_2=[-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}]^T$$

所以对应的矩阵 P 为：

$$P=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$$

可以验证协方差矩阵C的对角化

$$PCP^T=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\frac{6}{5} & \frac{4}{5}\\ \frac{4}{5} & \frac{6}{5}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}=\begin{bmatrix}2 & 0\\ 0 & \frac{2}{5}\end{bmatrix}$$

用 P 的第一行进行降维即可：

$$Y=\begin{bmatrix}\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}-1 & -1 & 0 & 2 & 0\\ -2 & 0 & 0 & 1 & 1\end{bmatrix}=\begin{bmatrix}-\frac{3}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \frac{3}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$$

d) 利用 LDA 降维，如下：

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

均值向量为：

$$u_1 = [-\frac{2}{3} \ -\frac{2}{3}]^T, u_2 = [1 \ \ 1]^T$$

则类内的散度矩阵为：

$$S_w = \Sigma_0 + \Sigma_1 = \sum_{x \in X_0} (x - u_0)(x - u_0)^T + \sum_{x \in X_1} (x - u_1)(x - u_1)^T$$

带入数据后

$$S_w = \begin{bmatrix} \frac{8}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{8}{3} \end{bmatrix}$$

类间的散度矩阵为：

$$s_b = (u_0 - u_1)(u_0 - u_1)^T$$

带入数据后有

$$S_b = \begin{bmatrix} \frac{25}{9} & \frac{25}{9} \\ \frac{25}{9} & \frac{25}{9} \end{bmatrix}$$

类内散度矩阵的逆矩阵为：

$$S_w^{-1} = \begin{bmatrix} \frac{2}{5} & -\frac{1}{10} \\ -\frac{1}{10} & \frac{2}{5} \end{bmatrix}$$

则$S_w^{-1} S_b$为：

$$S_w^{-1} S_b = \begin{bmatrix} \frac{5}{6} & \frac{5}{6} \\ \frac{5}{6} & \frac{5}{6} \end{bmatrix}$$

其特征值和特征向量为：

$$\lambda_1 = \frac{5}{3}, \lambda_2 = 0$$

$$c_1 = \frac{1}{\sqrt{2}}[1,1]^T, \ \ c_2 = \frac{1}{\sqrt{2}}[-1,1]^T$$

则投影矩阵w为：

$$w = \frac{1}{\sqrt{2}}[1,1]^T$$

降维后的数据为:

$$z = w^T X = \frac{1}{\sqrt{2}} * [1,1] * \begin{bmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \frac{3}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

LDA vs PCA

- **相同点：**
  - **两者均可以对数据进行降维。**
  - **两者在降维时均使用了矩阵特征分解的思想。**
  - **两者都假设数据符合高斯分布。**
- **不同点：**
  - **LDA是有监督的降维方法，而PCA是无监督的降维方法**
  - **LDA降维最多降到类别数k-1的维数，而PCA没有这个限制。**
  - **LDA除了可以用于降维，还可以用于分类。**
  - **LDA选择分类性能最好的投影方向，而PCA选择样本点投影具有最大方差的方向。**

Problem 4. （20 分）如图2是一个有 d 维输入，l 维输出和 q 维单隐藏层的多层感知机。该感知机的输出激活函数为 Sigmoid 函数，误差函数为均方误差 (MSE):，其中 $\hat{y}$ 为输出真值。请推导该感知机使用后向传播算法 (Back Propagation) 更新参数 w_{hj} 和 v_{ih} 一次的过程。

Solution: 对于输出层：

$$y_j = f(\text{net}_j), j = 1, 2, ..., l$$

$$\text{net}_j = \sum_{h=1}^{q} w_{hj} b_h, j = 1, 2, ..., l$$

对于隐藏层：

$$b_h = f(\text{net}_h), h = 1, 2, ..., q$$

$$\text{net}_h = \sum_{i=1}^{d} v_{ih} x_i, h = 1, 2, ..., q$$

损失函数 Sigmoid: $f(x) = \frac{1}{1+e^{-x}}$，$\frac{\partial f}{\partial x} = f(x)(1 - f(x))$

输出误差 E 定义：$E = \frac{1}{l}\sum_{j=1}^{l}(\hat{y}_j - y_j)^2$，将以上误差定义式展开至隐层：

$$\begin{aligned} E &= \frac{1}{l}\sum_{j=1}^{l}[\hat{y}_j - f(\text{net}_j)]^2 \\ &= \frac{1}{l}\sum_{j=1}^{l}[\hat{y}_j - f(\sum_{h=1}^{q} w_{hj} b_h)]^2 \end{aligned}$$

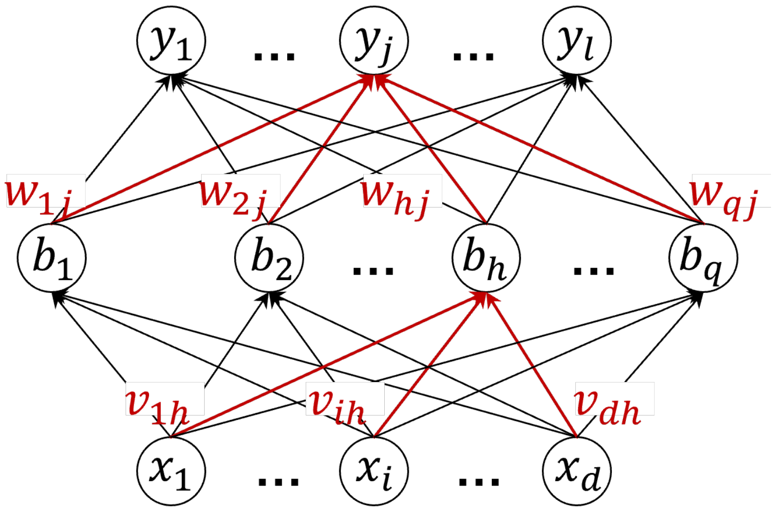


图 2: 单隐藏层感知机。

进一步展开至输入层:

$$
\begin{aligned}
E &= \frac{1}{l}\sum_{j=1}^{l}\{\hat{y}_j - f[\sum_{h=1}^{q} w_{hj} f(\text{net}_h)]\}^2 \\
&= \frac{1}{l}\sum_{j=1}^{l}\{\hat{y}_j - f[\sum_{h=1}^{q} w_{hj} f(\sum_{i=1}^{d} v_{ih}x_i)]\}^2
\end{aligned}
$$

由后向传播算法:

$$
\begin{aligned}
\Delta w_{hj} &= -\eta\frac{\partial E}{\partial w_{hj}} = -\eta\frac{\partial E}{\partial \text{net}_j}\frac{\partial \text{net}_j}{\partial w_{hj}}, \\
\Delta v_{ih} &= -\eta\frac{\partial E}{\partial v_{ih}} = -\eta\frac{\partial E}{\partial \text{net}_h}\frac{\partial \text{net}_h}{\partial v_{ih}},
\end{aligned}
$$

其中 η 为损失函数更新的比例系数。

$$
\begin{aligned}
\frac{\partial E}{\partial \text{net}_j} &= \frac{2}{l}[f(\text{net}_j) - \hat{y}_j]\cdot f(\text{net}_j)\cdot[1 - f(\text{net}_j)] \\
&= \frac{2}{l}[f(\sum_{h=1}^{q} w_{hj}b_h) - \hat{y}_j]\cdot f(\sum_{h=1}^{q} w_{hj}b_h)\cdot[1 - f(\sum_{h=1}^{q} w_{hj}b_h)], \\
\frac{\partial \text{net}_j}{\partial w_{hj}} &= b_h,
\end{aligned}
$$

$$\frac{\partial E}{\partial \text{net}_h} = \frac{2}{l}\sum_{j=1}^{l}\{f[\sum_{h=1}^{q} w_{hj} f(\text{net}_h)] - \hat{y}_j\} \cdot f[\sum_{h=1}^{q} w_{hj} f(\text{net}_h)] \cdot \{1 - f[\sum_{h=1}^{q} w_{hj} f(\text{net}_h)]\}$$
$$\cdot w_{hj} \cdot f(\text{net}_h)$$
$$\frac{\partial \text{net}_h}{\partial v_{ih}} = x_i,$$

更新过程:

$$w_{hj} = w_{hj} + \Delta w_{hj}$$
$$v_{ih} = v_{ih} + \Delta v_{ih}$$

Problem 5. （20 分）在一个图像处理深度神经网络中，第一层是 CNN 层，X 是输入图片，有 RGB 三个通道，分别是如下 X_1, X_2, X_3 三个矩阵:

$$X_1 = \begin{bmatrix} 5 & 0 & 3 & 5 & 1 \\ 2 & 8 & 8 & 7 & 4 \\ 3 & 3 & 3 & 3 & 3 \\ 7 & 7 & 2 & 9 & 8 \end{bmatrix}, X_2 = \begin{bmatrix} 0 & 1 & 6 & 2 & 8 \\ 4 & 1 & 2 & 9 & 3 \\ 7 & 0 & 7 & 8 & 8 \\ 1 & 5 & 3 & 4 & 1 \end{bmatrix}, X_3 = \begin{bmatrix} 6 & 1 & 8 & 9 & 3 \\ 7 & 0 & 6 & 6 & 9 \\ 9 & 9 & 5 & 1 & 7 \\ 5 & 6 & 2 & 5 & 3 \end{bmatrix}.$$

K 是卷积核，有三个通道构成，分别是如下 K_1, K_2, K_3 三个矩阵:

$$K_1 = \begin{bmatrix} 7 & 0 & 6 \\ 1 & 1 & 1 \\ 9 & 3 & 0 \end{bmatrix}, K_2 = \begin{bmatrix} 5 & 4 & 8 \\ 1 & 3 & 9 \\ 0 & 5 & 8 \end{bmatrix}, K_3 = \begin{bmatrix} 4 & 4 & 1 \\ 5 & 8 & 2 \\ 7 & 8 & 1 \end{bmatrix}.$$

(a) 计算图片 X 经过卷积核 K 生成的输出结果，步长为 1，没有填充。

(b) 卷积层后是二维 MaxPool 层，kernel size 是 2x2，步长为 1，没有填充，写出 MaxPool 层的输出。

Solution: 答案如图3所示。

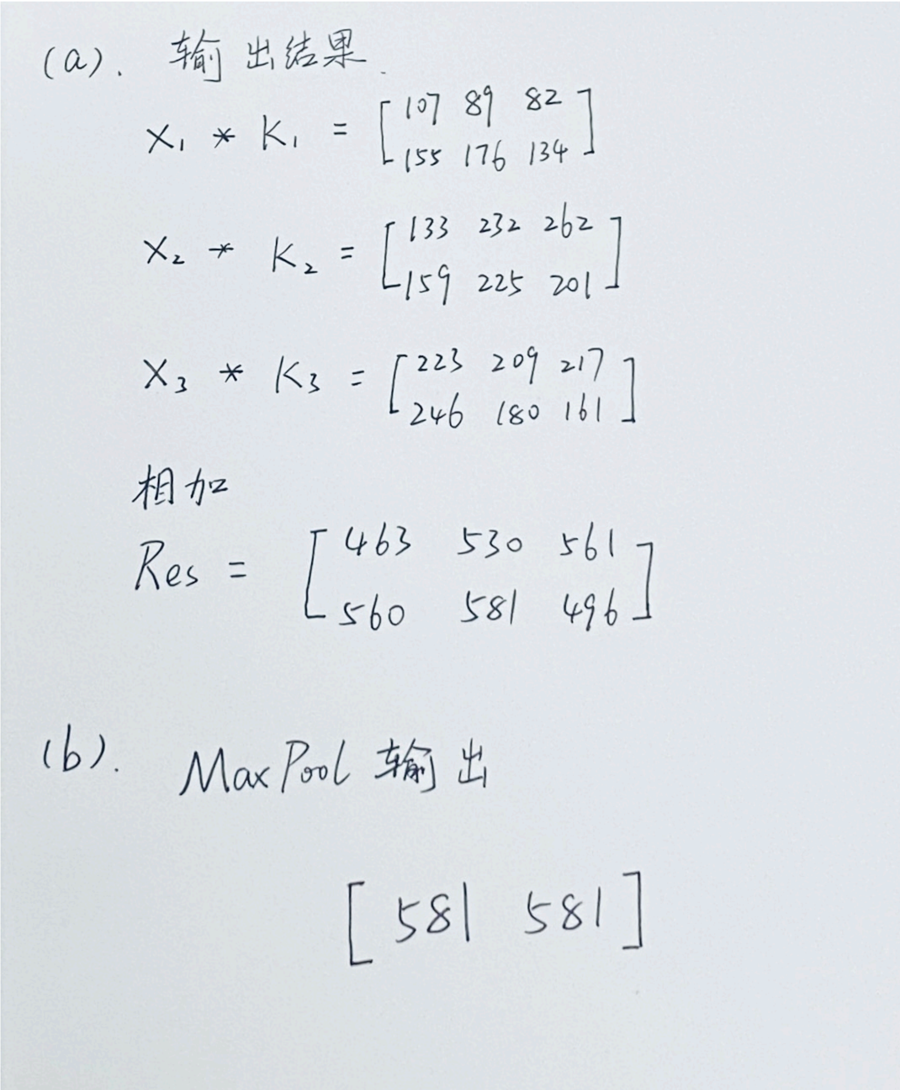

(a). 输出结果
X1 * K1 = [107 89 82; 155 176 134]
X2 * K2 = [133 232 262; 159 225 201]
X3 * K3 = [223 209 217; 246 180 161]
相加
Res = [463 530 561; 560 581 496]
(b). MaxPool 输出
[581 581]

图 3: CNN 答案