Homework #2

Due: 2025-11-16 23:59 | 7 Problems, 100 Pts

Name: 赵凌哲, ID: 2300015881

Problem 1 (20'). Consider a cube C with side length 1 in d-dimensional space.

(1) (4') Write down the radius of a ball B whose volume is equal to the cube's. You don't need to prove your result.

(2) (16') When the centers of the cube C and the ball B are both at the origin, calculate the volume of their intersection when $d \to +\infty$.

[Hint: Try to calculate $\mathbb{E}_{X\sim C}[\|X\|_2^2]$ and $Var_{X\sim C}[\|X\|_2^2]$. Then use Chebyshev's Inequality to analyze the concentration of $\|X\|_2^2$. You may use the Stirling's approximation

$$\lim_{n\to+\infty} \frac{\Gamma(n+1)}{\sqrt{2\pi n}(n/\mathrm{e})^n} = 1.$$

]

◀

Answer.

(1)

$$r = \left(\frac{\Gamma(\frac{d}{2}+1)}{\pi^{\frac{d}{2}}}\right)^{\frac{1}{d}}$$

(2) 首先进行计算：

$$X \sim Unif(\left[-\frac{1}{2}, \frac{1}{2}\right]^d)$$

$$\mathbb{E}X_i^2 = Var(X_i) = \frac{1}{12}$$

$$\mathbb{E}_{X\sim C}[\|X\|_2^2] = \sum_{i=1}^{d} \mathbb{E}X_i^2 = \frac{d}{12}$$

$$Var(X_i^2) = \mathbb{E}X_i^4 - (\mathbb{E}X_i^2)^2 = \frac{1}{180}$$

$$Var_{X\sim C}[\|X\|_2^2] = \sum_{i=1}^{d} Var(X_i^2) = \frac{d}{180}$$

又因为根据 Stirling's approximation：

$$r^2 = \left(\frac{\Gamma(\frac{d}{2}+1)}{\pi^{\frac{d}{2}}}\right)^{\frac{2}{d}} \sim \frac{d}{2e\pi}, \quad when\ d \to +\infty$$

容易知道，当 $d \to +\infty$ 时：

$$\mathbb{E}_{X\sim C}[\|X\|_2^2] > r^2$$
$$\mathbb{E}_{X\sim C}[\|X\|_2^2] - r^2 = O(d)$$

从而由 Chebyshev's Inequality 可以知道：

$$\mathbb{P}\big(\|X\|_2^2 \le r_d^2\big) \le \frac{Var_{X\sim C}[\|X\|_2^2]}{\big(\mathbb{E}_{X\sim C}[\|X\|_2^2] - r_d^2\big)^2} = O\left(\frac{d}{d^2}\right) \to 0, \quad when\ d \to +\infty$$

上述概率即为 C 的体积落在 B 中的比例，又因为 C 的体积恒为 1，从而有：

$$V(C \cap B) \to 0, \quad when\ d \to +\infty$$

◁

Problem 2 (16'). You select a PE class this term, so you need to complete a total of 85km of extracurricular exercise. There are only n days before the deadline, but you still have m kilometers remaining. To simplify your work, you can run at most 10km every day, but there is no lower bound. You are mindless when running, so the distance you run every day is uniformly random. In other words, the number of kilometers you run every day is a real number uniformly selected from range $[0, 10]$. Prove that, the probability that you can complete m kilometers in n days, i.e., the probability that your total distance in n days is greater than or equal to m kilometers is

$$1 - \sum_{i=0}^{\min(\lfloor m/10 \rfloor, n)} \frac{(-1)^i}{i!(n-i)!} \left(\frac{m}{10} - i\right)^n.$$

◀

Answer. 将每日里程和 m 都除以 10，从而归一化：

$$V_i \sim \mathrm{Unif}([0,1])$$
$$T_n = V_1 + \cdots + V_n$$

我们即求 $P\left(T_n \ge \frac{m}{10}\right) = 1 - F_{T_n}\left(\frac{m}{10}\right)$，其中 $F_{T_n}(x) = \mathbb{P}(T_n \le x)$.

(1) 当 $0 < x < 1$ 时：

$$F_{T_n}(x) = \frac{x^n}{n!}$$

这就是一个 n 维单纯形的体积，在第一次作业中已求出表达式.

(2) 当 $x \ge 1$ 时：

直观理解是我们需要把一个 n 维单纯形中，存在大于 1 分量的部分体积除去，因此定义：

$$S(x) = \{v \in \mathbb{R}_+^n : \ v_1 + \cdots + v_n \le x\}$$
$$A_i = \{v \in S(x) : \ v_i > 1\}, \quad i = 1, \ldots, n$$

由容斥原理得到：

$$\mathrm{Vol}\Big(S(x) \setminus \bigcup_{i=1}^{n} A_i\Big) = \mathrm{Vol}(S(x)) - \sum_i \mathrm{Vol}(A_i) + \sum_{i<j} \mathrm{Vol}(A_i \cap A_j) - \cdots .$$

当 $I \subset \{1, \ldots, n\}$ 且 $|I| = i$ 时：

$$\mathrm{Vol}\Big(\bigcap_{j \in I} A_j\Big) = \mathrm{Vol}\big(S(x-i)\big) = \frac{(x-i)^n}{n!}$$

上式直观理解是把每个大于 1 的分量减去 1 后仍然保证大于 0.
又因为 $|I| = i$ 的不同取法有 $\binom{n}{i}$ 种，代入上式知：

$$F_{T_n}(x) = \sum_{i=0}^{\lfloor x \rfloor} (-1)^i \binom{n}{i} \frac{(x-i)^n}{n!}$$

注意到上式在 $0 < x < 1$ 时同样成立，此时只有 $i = 0$ 一项.

因此我们可以统一表达式为上式，从而得到目标式子：

$$\begin{aligned} P\left(T_n \geq \frac{m}{10}\right) &= 1 - F_{T_n}\left(\frac{m}{10}\right) \\ &= 1 - \sum_{i=0}^{\lfloor m/10 \rfloor} (-1)^i \binom{n}{i} \frac{\left(\frac{m}{10} - i\right)^n}{n!} \\ &= 1 - \sum_{i=0}^{\min(\lfloor m/10 \rfloor, n)} \frac{(-1)^i}{i!(n-i)!} \left(\frac{m}{10} - i\right)^n \end{aligned}$$

把求和上限改为 $\min(\lfloor m/10 \rfloor, n)$ 的依据是 $i > n$ 时每一项都为 0. ◁

Problem 3 (10'). If A is square, show that $AA^\top$ and $A^\top A$ are similar.

[Hint: Use the SVD of A.] ◀

Answer. 考虑矩阵 A 的 SVD 结果：

$$A = U\Sigma V^\top, \quad U, V \textit{ are orthogonal}$$

从而有：

$$\begin{aligned} AA^\top &= U\Sigma V^\top V \Sigma U^\top = U\Sigma^2 U^\top \\ A^\top A &= V\Sigma U^\top U \Sigma V^\top = V\Sigma^2 V^\top \end{aligned}$$

我们只需要取 $Q = UV^\top$，由于 U, V 的正交性，Q 是可逆的，从而：

$$Q(A^\top A)Q^{-1} = UV^\top(V\Sigma^2 V^\top)VU^\top = U\Sigma^2 U^\top = AA^\top$$

因此有 $AA^\top$ and $A^\top A$ are similar. ◁

Problem 4 (12'). Calculate the SVD of matrix A, where

$$A = \begin{pmatrix} -4 & -6 \\ 3 & -8 \end{pmatrix}.$$

You need to calculate it in two different ways.

[Hint: Use the definition of singular vectors, or consider $A^\top A$.] ◀

Answer.

(1) 按特征值定义来做：

$$\max_{||v||=1} ||Av||^2 = v^\top A^\top Av = 25v_1^2 + 100v_2^2$$

注意到取 $v_1 = (0,1)^\top$ 即可使之最大化，再有：

$$\max_{||v||=1,\ v\perp v_1} ||Av||^2 = v^\top A^\top Av = 25v_1^2 + 100v_2^2$$

从而取 $v_2 = (1,0)^\top$ 使之最大化，综上可知：

$$V = \begin{bmatrix} v_1 \ v_2 \end{bmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \Sigma = \mathrm{diag}(10, 5)$$

$$U = AV\Sigma^{-1} = \begin{pmatrix} -3/5 & -4/5 \\ -4/5 & 3/5 \end{pmatrix}$$

从而得到 A 的 SVD，$A = U\Sigma V^\top$.

(2) 直接计算 $A^\top A$ 的特征值：

$$A^\top A = \begin{pmatrix} 25 & 0 \\ 0 & 100 \end{pmatrix}$$

为对角矩阵，从而奇异值从大到小为：

$$\sigma_1 = 10, \quad \sigma_2 = 5$$

取 $v_1 = (0,1)^\top$，$v_2 = (1,0)^\top$ 两特征向量，对应两个奇异值从大到小的顺序，从而可以得到 V，Σ，后续步骤同方法 (1)：

$$V = \begin{bmatrix} v_1 \ v_2 \end{bmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \Sigma = \mathrm{diag}(10, 5)$$

$$U = AV\Sigma^{-1} = \begin{pmatrix} -3/5 & -4/5 \\ -4/5 & 3/5 \end{pmatrix}$$

从而得到 A 的 SVD，$A = U\Sigma V^\top$.

◁

Problem 5 (16'). Let $A = \sum_{i=1}^{r} \sigma_i u_i v_i^\top$ be the SVD of a rank-r matrix A, let $A_k := \sum_{i=1}^{k} \sigma_i u_i v_i^\top$ be the rank-k approximation of A for some $k < r$.

(1) (4') Express the following quantities in terms of the singular values $\{\sigma_i, 1 \le i \le r\}$. You don't need to prove your result.

(a) (2') $\|A_k\|_F^2$.

(b) (2') $\|A_k\|_2^2$.

(2) (3') Prove that, $\sigma_k \le \frac{\|A\|_F}{\sqrt{k}}$ for $k = 1, 2, \cdots, r$.

(3) (9') Suppose $A \in \mathbb{R}^{m\times m}$. Let $B = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^\top$.

(a) (6') Show that $BAx = x$ for all x in the span of the right singular vectors of A. For this reason B is sometimes called the pseudo inverse of A and can play the role of A^{-1} in many applications.

(b) (3') Does $BAx = x$ always hold for all $x \in \mathbb{R}^m$? Prove or give a counterexample.

◀

Answer.

(1) (a) $\|A_k\|_F^2 = \sum_{i=1}^{k} \sigma_i^2$

(b) $\|A_k\|_2^2 = \sigma_1^2$

(2)

$$\|A\|_F^2 \ge \|A_k\|_F^2 = \sum_{i=1}^{k} \sigma_i^2 \ge k \cdot \sigma_k^2$$

$$\Longrightarrow \sigma_k \le \frac{\|A\|_F}{\sqrt{k}}, \quad for\ k = 1, 2, \cdots, r$$

(3) (a) 若 $x \in \mathrm{span}\{v_1, \ldots, v_r\}$，记 $x = \sum_{i=1}^{r} \alpha_i v_i$，则有：

$$Ax = \sum_{i=1}^{r} \alpha_i \sigma_i u_i,$$

$$BAx = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^\top \sum_{j=1}^{r} \alpha_j \sigma_j u_j = \sum_{i=1}^{r} \alpha_i v_i = x.$$

故对右奇异向量张成的子空间内恒有 $BAx = x$.

(b) 不一定，当 A 不满秩，只需要取：

$$x \perp \mathrm{span}\{v_1, \ldots, v_r\}$$

$$Ax = 0 \Rightarrow BAx = 0 \ne x$$

例如取 $A = \mathrm{diag}(1, 0) \in \mathbb{R}^{2\times 2}$，$x = (0, 1)^\top$.

◁

Problem 6 (16'). In class, we introduced the best-fit subspace for a set of points. We extend this definition to probability densities instead of a set of points. If P is a probability density in $\mathbb{R}^d$, then we define the best-fit k-dimensional subspace ($k \le d$)

$$V_k = \arg\max_{V, \dim(V)=k} \mathbb{E}_{X\sim P}(|\mathrm{proj}(X,V)|^2),$$

where $\mathrm{proj}(X,V)$ is the orthogonal projection of X onto V.

(1) (6') Find out the best-fit 1-dimensional subspace (i.e., a best-fit line through the origin) for probability density $\mathcal{N}\left(0,0,1,2,\frac{1}{2}\right)$.

[Hint: Suppose random variable $(X_1,X_2) \sim \mathcal{N}\left(0,0,1,2,\frac{1}{2}\right)$, then we have

$$X_1 \sim \mathcal{N}(0,1),\ X_2 \sim \mathcal{N}(0,2),\ \rho := \frac{\mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)}{\sqrt{Var(X_1)Var(X_2)}} = \frac{1}{2}.$$ *]*

(2) (10') Prove that, for d-dimensional Gaussian distribution $\mathcal{N}(\mu, I_d)$, a k-dimensional subspace is a best-fit subspace if and only if it contains μ.

[Hint: Suppose random variable $(X_1,\cdots,X_d) \sim \mathcal{N}(\mu, I_d)$, then we have $X_i \sim \mathcal{N}(\mu_i, 1)$ and $X_1,\cdots,X_d$ are independent random variables.]

◀

Answer.

(1) 由题给条件知协方差矩阵：

$$\Sigma = \begin{pmatrix} 1 & \rho\sqrt{1\cdot 2} \\ \rho\sqrt{1\cdot 2} & 2 \end{pmatrix} = \begin{pmatrix} 1 & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 2 \end{pmatrix}$$

因为均值为原点，只需要求出协方差矩阵的最大特征值和对应的特征向量，对应的特征向量方向即为 the best-fit 1-dimensional subspace：

$$\det(\Sigma - \lambda I) = \lambda^2 - 3\lambda + \tfrac{3}{2} = 0 \Rightarrow \lambda_{\max} = \frac{3+\sqrt{3}}{2}$$
$$\Longrightarrow v_{\max} \propto \begin{pmatrix} \lambda_{\max} - 2 \\ \frac{\sqrt{2}}{2} \end{pmatrix} = \begin{pmatrix} \frac{-1+\sqrt{3}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$$

此即为 the best-fit 1-dimensional subspace 的方向（也可以进行单位化）.

(2) 定义：

$$X = \mu + Z, \quad Z \sim N(0, I_d), \quad V_k = \arg\max_{\dim V=k} \mathbb{E}\big(|\,\mathrm{proj}(X,V)|^2\big)$$

从而根据 $X=\mu+Z$ 我们有：

$$\mathbb{E}|\operatorname{proj}(X,V)|^2 = \underbrace{|\operatorname{proj}(\mu,V)|^2}_{\text{依赖}\mu,\,V} + \underbrace{\mathbb{E}|\operatorname{proj}(Z,V)|^2}_{=k,\ \text{不依赖}V},$$

第二项等于 k 是因为 Z_i 间相互独立，从而知道该均值展开后，只有平方项不为零，结果为投影矩阵的迹，也就是 k.

从而原问题等价于最大化 $|\operatorname{proj}(\mu,V)|^2$，其最大值为 $|\mu|^2$，当且仅当 $\mu\in V$ 时取到.

故结论成立：V 是最佳拟合的 k 维子空间 $\iff$ V 包含 μ.

◁

Problem 7 (10'). Read in a photochrome and convert to a matrix. Perform a singular value decomposition of the matrix. Reconstruct the photo using only 1,2,4,16 singular values.

(1) Print the reconstructed photo. How good is the quality of the reconstructed photo? Write down your observations.

(2) What percent of the Frobenius norm is captured in each case? Describe the way you calculate the captured Frobenius norm.

You need to include your original photochrome and all reconstructed photos in your answer and submit the code as an attachment.

[Hint: You may want to perform SVD on each channel.] ◀

Answer. 为了进行对比，分别选取色彩较为单一和较为复杂的两张图片，进行操作.

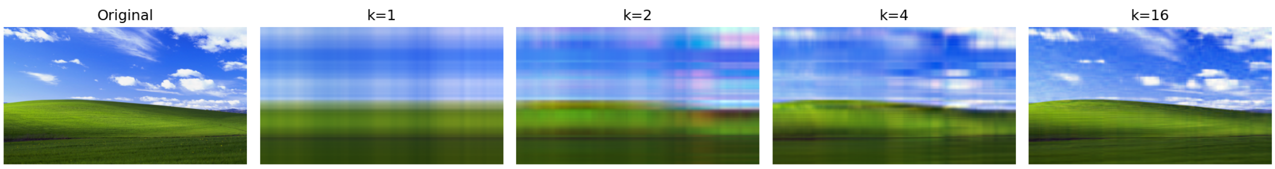

图 1: 某著名壁纸

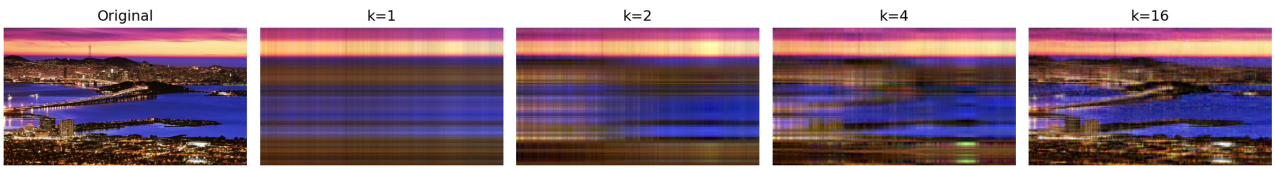

图 2: 某风景照

(1) 观察到，在 k=1 时，还原的图片为秩一（图中只存在完全水平和竖直的纹路）. k=2 时，出现了更多色块，颜色种类增加. k=4 则进一步增加颜色种类，图 1 已经较为接近原始图片.k=16 时，图 1 已经相当接近原始图片，图 2 也显示出原始图片的全部轮廓，但是仍不如图 1.

(2) 在原始代码中，我分别将每一个 channel 进行了 SVD 重建，在计算 Frobenius norm 时，我的方法为：

$$\text{ratio}(k) = \frac{\sum_{\text{通道}c} \sum_{i=1}^{k} \sigma_{c,i}^2}{\sum_{\text{通道}c} \sum_{i=1}^{r_c} \sigma_{c,i}^2}$$

结果如下表：

k 值	**图 1 占比**	**图 2 占比**
1	95.56%	82.19%
2	97.15%	85.73%
4	98.50%	88.38%
16	99.12%	92.63%

表 1: 不同 k 值下图像重建的 Frobenius 范数占比 (%)

从表中也可以明显看出，图 1 的重建效果明显好于图 2. 图 2 由于图中灯光等细节过多，重建效果始终较为一般.

另外对于 Frobenius norm 都达到 80% 至 90% 数量级的原因，也较容易解释：因为大部分能量来自于大块的均匀背景和整体亮度. 在这个任务中，我们没有必要进行均值处理，因此这个比例反映的不完全是对于色彩变化的捕捉. 即使如此，仍然可以看出图 1 和图 2 重建结果存在明显差异，说明 Frobenius norm 占比具有参考价值.

本题使用的代码参见附件的 *svd.py* . ◁