CS 3220: Prelim 2 Solutions

Instructor: Anil Damle

Name and netID:

Policies and details

This document constitutes the second preliminary exam for 3220. It contains 3 questions and you have one hour and thirty minutes to complete the exam. Please neatly write your solutions directly onto the exam booklet (scratch paper is available if necessary). All of the work on this exam must be your own and this exam is governed by the Cornell Academic Integrity Code. It is a violation of the Academic Integrity Code to look at any exam other than your own, utilize any references, or otherwise give or receive any unauthorized assistance. Please avoid discussing this exam with any students scheduled to take the exam at an alternative time.

This exam is constructed to assess your grasp of the material in the class over several levels of difficulty—some of the questions may be difficult. Please do your best to answer all of the questions and provide partial solutions if you have them, partial credit will be awarded as appropriate. If you have any questions during the exam please ask, I will not provide hints but if something is unclear I am happy to clarify.

Question 1

Let $Z \in \mathbb{R}^n$ be distributed as $\mathcal{N}(0, I)$. Any affine transform of the random variables Z denoted $X = AZ + b$ for $A \in \mathbb{R}^{m \times n}$ with $m \leq n$, linear independent rows, and $b \in \mathbb{R}^m$ follows a multivariate normal distribution.

(a) Derive the mean and covariance matrix for X, *i.e.* X is distributed as $\mathcal{N}(\mu, \Sigma)$—compute μ and Σ in terms of A and b.

(b) For any fixed matrix $V \in \mathbb{R}^{n \times k}$ with orthonormal columns show that

$$\mathbb{E}[\|VV^T Z\|_2^2] = k.$$

(c) Show that for any $t \geq 0$

$$\mathbb{P}\left\{|\|VV^T Z\|_2^2 - k| \geq t\sqrt{k}\right\} \leq \frac{2}{t^2}.$$

A fact you may use: A random variable distributed as $\chi^2(d)$, *i.e.* the distribution of $\sum_{i=1}^{d} Y_i^2$ where the Y_i are independent $\mathcal{N}(0,1)$ random variables, has expected value d and variance $2d$

Solution

(a) First, since $\mathbb{E}[AZ+b] = A\mathbb{E}[Z]+b$ and $\mathbb{E}[Z] = 0$ we conclude that $\mu = b$. Second, since $\mathbb{E}[X] = b$ we have that the covariance matrix is

$$\begin{aligned}
\Sigma &= \mathbb{E}[(X-b)(X-b)^T] \\
&= \mathbb{E}[(AZ)(AZ)^T] \\
&= A\mathbb{E}[ZZ^T]A^T \\
&= AA^T.
\end{aligned}$$

(b) Since $\|VV^TZ\|_2^2 = \|V^TZ\|_2^2$ and from the previous part we have that $V^TZ \sim \mathcal{N}(0, I)$, we can conclude that $\|V^TZ\|_2^2$ is distributed as a $\chi^2(k)$ random variable. From this observation we may immediately conclude that $\mathbb{E}[\|VV^TZ\|_2^2] = k$.

(c) By Chebyshev's inequality we have that for $a > 0$

$$\mathbb{P}\left\{|\|VV^TZ\|_2^2 - k| \geq a\right\} \leq \frac{2k}{a^2}.$$

Letting $a = t\sqrt{k}$ yields the desired result.

Question 2

Assume we are given a matrix $V \in \mathbb{R}^{d\times k}$ with orthonormal columns and $k < d$ and generate a set of n independent and identically sampled vectors $X_1, X_2, \ldots, X_n$ as $X_i = VZ + \mu$ where Z is distributed as $\mathcal{N}(0, I_k)$ and $V^T\mu = 0$. As usual, let $X \in \mathbb{R}^{d\times n}$ be the matrix whose columns are the random samples, *i.e.* $X(:, i) = X_i$, and let $\widehat{X}$ be the version of X whose rows have mean zero.[1]

(a) Assume we forget to remove the mean from the rows of X and accidentally compute the principle components as the eigenvectors of XX^T rather than $\widehat{X}\widehat{X}^T$. To explore what happens (besides the fact that we are not technically computing sample covariances) show that if $\|\mu\|_2 > 1$ then the eigenvector associated with the eigenvalue of largest magnitude of $\mathbb{E}[XX^T]$ is μ.

(b) Using some of the limit theorems and asymptotic behavior results that we have from class can you argue about what happens when we do not take the expectation but let $n \to \infty$? (It is fine to be informal here.) In other words, argue that in the same setting as above the eigenvector associated with XX^T "converges" to μ as $n \to \infty$. (You may assume $\|\mu\|_2 \gg 1$ if you like, but note that μ does not depend on n.)

Solution

(a) From the problem statement we have that $X = VZ + \mu\mathbf{1}^T$ where Z_i are i.i.d. $\mathcal{N}(0, I)$ and $\mathbf{1} \in \mathbb{R}^n$ is the vector of all ones. We then have that

$$\begin{aligned}
\mathbb{E}[XX^T] &= \mathbb{E}[(VZ + \mu\mathbf{1}^T)(VZ + \mu\mathbf{1}^T)^T] \\
&= \mathbb{E}[VZZ^TV^T + VZ\mathbf{1}\mu^T + \mu\mathbf{1}^TZ^TV^T + \mu\mathbf{1}^T\mathbf{1}\mu^T] \\
&= nVV^T + n\mu\mu^T
\end{aligned}$$

where we have used that $\mathbb{E}[ZZ^T] = nI$ and $\mathbb{E}[Z] = 0$. Since scaling a matrix doesn't change its eigenvectors we just need to show that the eigenvalue of largest magnitude associated with

$$\mathbb{E}[XX^T] = VV^T + \mu\mu^T$$

is in the direction of μ. There are several ways to show this, one is to observe that we may write (recall that the columns of V and μ are orthogonal)

$$\mathbb{E}[XX^T] = \begin{bmatrix} V & \frac{\mu}{\|\mu\|_2} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \|\mu\|_2^2 \end{bmatrix} \begin{bmatrix} V^T \\ \frac{\mu^T}{\|\mu\|_2} \end{bmatrix}.$$

This is an eigen-decomposition of $\mathbb{E}[XX^T]$ omitting zero eigenvalues. Since the eigenvalues are 1 and $\|\mu\|_2$, if $\|\mu\|_2 > 1$ we have that the eigenvalue associated with the largest magnitude eigenvalue is in the direction of μ.

[1] Recall that if we let $\bar{\mu} = \frac{1}{n}\sum_{i=1}^n X_i$ then $\widehat{X} = X - \bar{\mu}\mathbf{1}^T$ where $\mathbf{1} \in \mathbb{R}^n$ is a vector of all ones.

(b) Similar to the above, and taking advantage of the fact that eigenvectors are invariant to scaling we have that

$$\begin{aligned}\frac{1}{n}XX^T &= \frac{1}{n}(VZ+\mu\mathbf{1}^T)(VZ+\mu\mathbf{1}^T)^T \\ &= \frac{1}{n-1}VZZ^TV^T + \frac{1}{n}VZ\mathbf{1}\mu^T + \frac{1}{n}\mu\mathbf{1}^TZ^TV^T + \frac{1}{n}\mu\mathbf{1}^T\mathbf{1}\mu^T.\end{aligned}$$

We now observe that $\frac{1}{n}Z\mathbf{1} \to 0$ since it is the sample mean and $\frac{1}{n}ZZ^T \to I$. This says that as $n \to \infty$ we have

$$\frac{1}{n}XX^T \to VV^T + \mu\mu^T$$

and the result follows from the same argument as above.

Question 3

Say we want to use rejection sampling to generate samples from a random variable X with pdf f_X. Assume we are given a means to generate samples of the random variable Y with pdf f_Y and that $f_X(x)/f_Y(x) \leq c$ for all x. Furthermore, assume there exists a $\tilde{x}$ such that $f_X(\tilde{x})/f_Y(\tilde{x}) = c$; in other words, the upper bound is tight.

(a) If we also assume that $\gamma \leq f_X(x)/f_Y(x)$ for all x and let N be the random variable denoting the number of iterations of rejection sampling we have to run before accepting a sample. Show that if we use rejection sampling with the best possible upper bound on $f_X(x)/f_Y(x)$ we have that

$$\mathbb{P}[N > k] \leq \left(1 - \frac{\gamma}{c}\right)^k$$

(b) Now, lets say we incorrectly compute our upper bound on $f_X(\tilde{x})/f_Y(\tilde{x})$ and instead run rejection sampling with $\tilde{c} = 2c$. In other words, we draw a sample from Y and a sample u from a uniform$[0,1]$ distribution and only accept Y as a sample from X if

$$u \leq \frac{f_X(Y)}{\tilde{c}f_Y(Y).}$$

Discuss what, if any, impact this would have on the prior result (for example, does the behavior with respect to k change?).

Solution

(a) Observe that to see $N > k$ we must reject k samples. At each step we independently reject the sample independently with probability

$$1 - \frac{f_X(x)}{cf_Y(x)}.$$

Using our bounds, this means that for each sample the probability of rejection is bounded from above by $(1-\gamma/c)$. As the rejections are independent we conclude that the probably k samples are rejected is at most $(1-\gamma/c)^k$.

(b) Nothing really changes, other than that we can instead say

$$\mathbb{P}[N > k] \leq \left(1 - \frac{\gamma}{2c}\right)^k.$$

While this nominally effects the probabilities, structurally it does not seem to change much.