Quiz - 0

CS 419M: Introduction to Machine Learning
Spring 2021-22
Department of Computer Science and Engineering
IIT Bombay

April 20, 2022

1. Show that

$$\mathbb{E}[Y|X \leq 0] = \frac{1}{F_X(0)} \int_{-\infty}^{0} \mathbb{E}[Y|X]\, f_X(x)\, dx$$

Solution:

$$\begin{aligned}
\mathbb{E}[Y|X \leq 0] &= \int_{-\infty}^{\infty} y\, f(y|X \leq 0)\, dy \\
&= \int_{-\infty}^{\infty} y\, \frac{f(y, x \leq 0)}{F_X(0)}\, dy \\
&= \frac{1}{F_X(0)} \int_{-\infty}^{\infty} \int_{-\infty}^{0} y\, f(y|x)\, f_X(x)\, dx\, dy \\
&= \frac{1}{F_X(0)} \int_{-\infty}^{0} \mathbb{E}[Y|X]\, f_X(x)\, dx
\end{aligned}$$

2. A random variable X assumes only non-negative integer values. Then show that

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$$

Solution:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{\infty} k\,\mathbb{P}(X=k) \\
&= \sum_{k=1}^{\infty} k\,\mathbb{P}(X=k) \\
&= \sum_{k=1}^{\infty} k\left(\mathbb{P}(X\geq k)-\mathbb{P}(X\geq k+1)\right) \\
&= \Big(\mathbb{P}(X\geq 1)+2\mathbb{P}(X\geq 2)+\ldots\Big)-\Big(\mathbb{P}(X\geq 2)+2\mathbb{P}(X\geq 3)+\ldots\Big) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(X\geq k)
\end{aligned}$$

3. (X,Y) are i.i.d. standard normal random variables in $\mathbb{R}^2$. Find the joint distribution of polar coordinates (R,θ)., $R\geq 0, \theta\in[0,2\pi]$.

Solution:

$$X = R\cos\theta$$

$$Y = R\sin\theta$$

Let g denote the transformation on the vector (X,Y) to get (R,θ). Then we have

$$g^{-1}(r,\theta) = (x,y) = (h_1(r,\theta), h_2(r,\theta))$$

This gives the Jacobian

$$J = \begin{pmatrix} \frac{\partial h_1}{\partial r} & \frac{\partial h_1}{\partial \theta} \\ \frac{\partial h_2}{\partial r} & \frac{\partial h_2}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

$$\implies \det J = r$$

So, the joint PDF of (R,θ) will be

$$\begin{aligned}
f_{R,\theta}(r,\theta) &= f_{X,Y}(r\cos\theta, r\sin\theta)\ \det J \\
&= f_X(r\cos\theta)\ f_Y(r\sin\theta)\ r \\
&= \frac{1}{\sqrt{2\pi}}\,e^{\frac{-r^2\cos^2\theta}{2}}\ \frac{1}{\sqrt{2\pi}}\,e^{\frac{-r^2\sin^2\theta}{2}} \\
&= \frac{1}{2\pi}\,r\,e^{\frac{-r^2}{2}}
\end{aligned}$$

4. A hundred students have taken an exam consisting of 8 problems, and for each problem at least 65 of the students got the right answer. Show that there exist two students who collectively got everything right, in the sense that for each problem, at least one of the two got it right.

Solution: Expected score of a random pair of students is $8(1-0.35^2) = 7.02$. Which means some pair must have got score 8.

5. Let X be a Pois(λ) random variable, where λ is fixed but unknown. Let $\theta = e^{-3\lambda}$, and suppose that we are interested in estimating θ based on the data. Since X is what we observe, our estimator is a function of X, call it g(X). The bias of the estimator g(X) is defined to be $E(g(X)) - \theta$, i.e., how far off the estimate is on average; the estimator is unbiased if its bias is 0. Show that $g(X) = (-2)^X$ is an unbiased estimator for θ.

Solution: The estimator $g(X) = (-2)^X$ is unbiased since

$$\begin{aligned} E(-2)^X - \theta &= \sum_{k=0}^{\infty} (-2)^k \frac{\lambda^k}{k!} e^{-\lambda} - e^{-3\lambda} \\ &= e^{-\lambda} e^{-2\lambda} - e^{-3\lambda} = 0 \end{aligned}$$

6. Let $Z \sim \mathcal{N}(0,1)$ and Φ and ϕ be the CDF and PDF of Z respectively. Show that for any $t > 0$, $I(Z > t) \leq (Z/t)I(Z > t)$. Using this show that, $\Phi(t) \geq 1 - \phi(t)/t$

Solution: $I(Z > t) \leq (Z/t)I(Z > t)$ is true since if indicator is 0 then both sides are 0, and if indicator is 1 then $Z/t > 1$. So,

$$E(I(Z > t)) \leq \frac{1}{t} E(ZI(Z > t)) = \frac{1}{t} \int_{-\infty}^{\infty} zI(z > t)\phi(z)dz = \frac{1}{t} \int_{t}^{\infty} z\phi(z)dz$$

Using substitution $u = z^2/2$, we have

$$\int ze^{-z^2/2}dz = \int e^{-u}du = -e^{-u} + C = -e^{-z^2/2} + C$$

Thus,

$$P(Z > t) = E(I(Z > t)) \leq \phi(t)/t$$

which proved the required bound on $\Phi(t)$.

7. Find the solution of the following optimization problem:

$$J(\boldsymbol{R}) = \|\boldsymbol{A} - \boldsymbol{R}\boldsymbol{B}\|_F^2$$

where $\boldsymbol{A} \in \mathbb{R}^{n \times m}, \boldsymbol{B} \in \mathbb{R}^{n \times m}, \boldsymbol{R} \in \mathbb{R}^{n \times n}, m > n$ and $\boldsymbol{R}$ is constrained to be orthonormal i.e., $\boldsymbol{R}^\top \boldsymbol{R} = \boldsymbol{R}\boldsymbol{R}^\top = \boldsymbol{I}$. $\boldsymbol{A}$, $\boldsymbol{B}$ are known. $\|\cdot\|_F$ denotes the Frobenius norm. (Hint: You might want to use Singular Value Decomposition (SVD). The SVD of a matrix $B\,A^\top$ is given by $B\,A^\top = U\,D\,V^\top$ where U and V are orthogonal matrices and D is a diagonal matrix.)

Solution:

$$\begin{aligned}
\min \|\boldsymbol{A}-\boldsymbol{R}\boldsymbol{B}\|_F^2 &= \min \operatorname{Tr}\Big((\boldsymbol{A}-\boldsymbol{R}\boldsymbol{B})^\top(\boldsymbol{A}-\boldsymbol{R}\boldsymbol{B})\Big)\\
&= \min \operatorname{Tr}\Big((\boldsymbol{A}^\top-\boldsymbol{B}^\top\boldsymbol{R}^\top)(\boldsymbol{A}-\boldsymbol{R}\boldsymbol{B})\Big)\\
&= \min \operatorname{Tr}\Big(\boldsymbol{A}^\top\boldsymbol{A}-2\boldsymbol{A}^\top\boldsymbol{R}\boldsymbol{B}+\boldsymbol{B}^\top\boldsymbol{B}\Big)\\
&= \max \operatorname{Tr}\Big(\boldsymbol{A}^\top\boldsymbol{R}\boldsymbol{B}\Big)\\
&= \max \operatorname{Tr}\Big(\boldsymbol{R}\boldsymbol{B}\boldsymbol{A}^\top\Big) \qquad \text{since trace(PQ) = trace(QP)}
\end{aligned}$$

using Singular Value Decomposition on $\boldsymbol{B}\boldsymbol{A}^\top$, we have

$$\begin{aligned}
\boldsymbol{B}\boldsymbol{A}^\top &= \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top \quad \text{where} \quad U, V \text{ are orthogonal and } D \text{ is diagonal}\\
\implies \min \|\boldsymbol{A}-\boldsymbol{R}\boldsymbol{B}\|_F^2 &= \max \operatorname{Tr}\Big(\boldsymbol{R}\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top\Big)\\
&= \max \operatorname{Tr}\Big(\boldsymbol{V}^\top\boldsymbol{R}\boldsymbol{U}\boldsymbol{D}\Big)\\
&= \max \operatorname{Tr}\Big(\boldsymbol{Z}(\boldsymbol{R})\boldsymbol{D}\Big)\\
&= \max \sum_i z_{ii}d_{ii} \le \sum_i d_{ii} \quad (\because \boldsymbol{Z}(\boldsymbol{R})^\top\boldsymbol{Z}(\boldsymbol{R}) = \boldsymbol{I})
\end{aligned}$$

The maximum is achieved for $\boldsymbol{Z}(\boldsymbol{R}) = \boldsymbol{I}$, i.e.,

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{U}^\top$$

8. If A is a symmetric positive definite matrix, then prove that

$$\left|x^\top Ay\right|^2 \le (x^\top Ax)(y^\top Ay)$$

Solution: As A is symmetric and positive definite, it is diagonalizable and can be represented as

$$A = Q\Lambda Q^\top = (Q\sqrt{\Lambda})(\sqrt{\Lambda}Q^\top) = R^\top R$$

. Hence,

$$\begin{aligned}
\left|x^\top Ay\right|^2 &= \left|x^\top R^\top Ry\right|^2\\
&= \left|(Rx)^\top(Ry)\right|^2\\
&\le \|Rx\|^2\,\|Ry\|^2\\
&= \Big(x^\top R^\top Rx\Big)\Big(y^\top R^\top Ry\Big)\\
&= (x^\top Ax)(y^\top Ay)
\end{aligned}$$

This inequality is also called as **Generalized Schwarz Inequality**.

9. Consider a Euclidean Distance Matrix (EDM) $\mathbf{D} \in \mathbb{R}_+^{n\times n}$ defined over a set of n points $\{x_1, x_2, \ldots, x_n\}, x_i \in \mathbb{R}^k$ such that $D_{ij} = \left\|x_i - x_j\right\|^2$. Prove that $\mathbf{D}$ has a rank of atmost $k+2$.

Solution: Denote $\mathbf{X} = \left(x_1 \mid x_2 \mid \ \ldots \mid x_n\right), \mathbf{X} \in \mathbb{R}^{k\times n}$. Then $\mathbf{D}$ can be represented as:

$$\mathbf{D} = \operatorname{diag}(\mathbf{X}\mathbf{X}^\top)\,\mathbf{1}^\top + \mathbf{1}\operatorname{diag}(\mathbf{X}\mathbf{X}^\top)^\top - 2\,\mathbf{X}\mathbf{X}^\top$$

Hence

$$\operatorname{rank}(\mathbf{D}) \leq \operatorname{rank}\big(\operatorname{diag}(\mathbf{X}\mathbf{X}^\top)\,\mathbf{1}^\top\big) + \operatorname{rank}\big(\mathbf{1}\operatorname{diag}(\mathbf{X}\mathbf{X}^\top)^\top\big) + \operatorname{rank}\big(2\,\mathbf{X}\mathbf{X}^\top\big) = 1 + 1 + k = k + 2$$