CSEP546 : Machine Learning: Homework 0

Karuna Sagar Krishna

January 21, 2025

Question 1

Final Answer: The probability that the patient has the disease given that the test result is positive is 0.98%

Let X be a random variable representing the outcome of the test result (positive/true or negative/false) and H be a random variable representing if patient has the disease (has disease/true or doesn't have disease/false). We are asked to find $P(H = true|X = true)$ i.e. the probability that patient has the disease given that the test result is positive.

We know $P(X = true|H = true) = 0.99$ and $P(X = false|H = false) = 0.99$ i.e. the test is 99% accurate. Also, we know that $P(H = true) = 1/10,000 = 0.0001$ i.e. the disease is rare. This means:

$$\begin{aligned} P(X = false|H = true) &= 1 - P(X = true|H = true) \\ &= 1 - 0.99 \\ &= 0.01 \end{aligned}$$

$$\begin{aligned} P(X = true|H = false) &= 1 - P(X = false|H = false) \\ &= 1 - 0.99 \\ &= 0.01 \end{aligned}$$

We also need $P(X = true)$ to compute $P(H = true|X = true)$:

$$\begin{aligned} P(X = true) &= P(X = true, H = true) + P(X = true, H = false) \\ &= P(X = true|H = true)P(H = true) + P(X = true|H = false)P(H = false) \\ &= (0.99 \times 0.0001 + 0.01 \times (1 - 0.0001)) \\ &= 0.010098 \end{aligned}$$

So, we can compute $P(H = true, X = true)$ as:

$$
\begin{aligned}
P(H = true|X = true) &= \frac{P(H = true, X = true)}{P(X = true)} \\
&= \frac{P(X = true|H = true)P(H = true)}{P(X = true)} \\
&= \frac{0.99 \times 0.0001}{0.010098} \\
&= 0.00980392156
\end{aligned}
$$

Question 2a

By applying distributive property and linearity of expectations, we get:

$$
\begin{aligned}
Cov(X, Y) &= E[(X - E(X))(Y - E(Y))] \\
&= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\
&= E[XY] - E[XE(Y)] - E[YE(X)] + E[E(X)E(Y)] \\
&= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] \\
&= E[XY] - E[X]E[Y]
\end{aligned}
$$

Assuming X and Y are discrete random variables. Also, say $Z = XY$ is a random variable. Using law of total expectations, we compute $E(XY)$ and $E(Y)$ as:

$$
\begin{aligned}
E(XY) &= E(Z) \\
&= E_X(E(Z|X)) \\
&= \sum_x E(Z|X = x)P(X = x) \\
&= \sum_x E(XY|X = x)P(X = x) \\
&= \sum_x E(xY|X = x)P(X = x) \\
&= \sum_x xE(Y|X = x)P(X = x) \\
&= \sum_x x^2 P(X = x) \\
&= E(X^2)
\end{aligned}
$$

$$
\begin{aligned}
E(Y) &= E_X(E(Y|X)) \\
&= \sum_x E(Y|X=x)P(X=x) \\
&= \sum_x xP(X=x) \\
&= E(X)
\end{aligned}
$$

So, plugging $E(XY)$ and $E(Y)$ back into $Cov(X,Y)$, we get:

$$
\begin{aligned}
Cov(X,Y) &= E[XY] - E[X]E[Y] \\
&= E[X^2] - E[X]E[X] \\
&= E[X^2] - E[X]^2
\end{aligned}
$$

Finally, we have:

$$
\begin{aligned}
E[(X-E(X))^2] &= E[X^2 - 2XE(X) + E(X)^2] \\
&= E[X^2] - E[2XE(X)] + E[E(X)^2] \\
&= E[X^2] - 2E[X]E(X) + E(X)^2 \\
&= E[X^2] - E(X)^2 \\
&= Cov(X,Y)
\end{aligned}
$$

Question 2b

If X and Y are independent, then $P(X,Y) = P(X)P(Y)$. Assuming X and Y are discrete random variables, $E(XY)$ can be rewritten as:

$$
\begin{aligned}
E(XY) &= \sum_{x,y} xyP(X=x, Y=y) \\
&= \sum_{x,y} xyP(X=x)P(Y=y) \\
&= \sum_x \sum_y xyP(X=x)P(Y=y) \\
&= \sum_x xP(X=x) \sum_y yP(Y=y) \\
&= E(X)E(Y)
\end{aligned}
$$

So, using $Cov(X,Y) = E(XY) - E(X)E(Y)$ from above, we get:

$$
\begin{aligned}
Cov(X,Y) &= E(XY) - E(X)E(Y) \\
&= E(X)E(Y) - E(X)E(Y) \\
&= 0
\end{aligned}
$$

Question 3a

Question 3b

Question 4a

$\mathcal{N}(0,1) \implies \mu = 0, \sigma^2 = 1$. Lets say $Y = aX_1 + b$, then we want $E(Y) = \mu = 0$ and $Var(Y) = \sigma^2 = 1$.

Using linearity of expectations i.e. $E(aX + b) = aE(X) + b$, we get:

$$\begin{aligned} E(Y) &= E(aX_1 + b) \\ &= aE(X_1) + b \\ &= a\mu + b \\ &= 0 \end{aligned}$$

Using $Var(X) = E[X^2] - E(X)^2$ and linearity of expectations, we get:

$$\begin{aligned} Var(Y) &= E[(Y - E(Y))^2] \\ &= E[(aX_1 + b - E(aX_1 + b))^2] \\ &= E[(aX_1 + b - aE(X_1) - b)^2] \\ &= E[(aX_1 - aE(X_1))^2] \\ &= E[(aX_1)^2 + (aE(X_1))^2 - 2a^2X_1E(X_1)] \\ &= E[a^2X_1^2 + a^2E(X_1)^2 - 2a^2X_1E(X_1)] \\ &= E[a^2X_1^2] + E[a^2E(X_1)^2] - E[2a^2X_1E(X_1)] \\ &= a^2E[X_1^2] + a^2E(X_1)^2 - 2a^2E[X_1]E(X_1) \\ &= a^2E[X_1^2] + a^2E(X_1)^2 - 2a^2E[X_1]^2 \\ &= a^2E[X_1^2] - a^2E[X_1]^2 \\ &= a^2[E(X_1^2) - E(X_1)^2) \\ &= a^2Var(X_1) \\ &= a^2\sigma^2 \\ &= 1 \end{aligned}$$

This gives us two equations that we can solve to get a and b:

$$a\mu + b = 0 \tag{1}$$

$$a^2\sigma^2 = 1 \tag{2}$$

From (2), we get: $a = \pm\frac{1}{\sigma}$. Plugging this into (1), we get: $b = \mp\frac{\mu}{\sigma}$

Question 4b

Both X_1 and X_2 are sampled from $\mathcal{N}(\mu, \sigma^2)$, so $E(X_1) = E(X_2) = \mu$ and $Var(X_1) = Var(X_2) = \sigma^2$. Using linearity of expectations:

$$\begin{aligned} E(X_1 + 2X_2) &= E(X_1) + 2E(X_2) \\ &= \mu + 2\mu \\ &= 3\mu \end{aligned}$$

Since X_1 and X_2 are independent, $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$. Further X_1 and $2X_2$ are also independent. And from previous question, we have $Var(aX_1) = a^2 Var(X_1)$. Using all of these, we get:

$$\begin{aligned} Var(X_1 + 2X_2) &= Var(X_1) + Var(2X_2) \\ &= Var(X_1) + 4Var(X_2) \\ &= \sigma^2 + 4\sigma^2 \\ &= 5\sigma^2 \end{aligned}$$

Question 4c

Applying linearity of expectations and using $E(X_i) = \mu$, we get:

$$\begin{aligned} E(\sqrt{n}(\widehat{\mu}_n - \mu)) &= \sqrt{n}[E(\widehat{\mu}_n) - E(\mu)] \\ &= \sqrt{n}[E(\frac{1}{n}\sum X_i) - \mu] \\ &= \sqrt{n}[\frac{1}{n}E(\sum X_i) - \mu] \\ &= \sqrt{n}[\frac{1}{n}(\sum E(X_i)) - \mu] \\ &= \sqrt{n}[\frac{1}{n}(n\mu) - \mu] \\ &= 0 \end{aligned}$$

Since X_i are independent, $Var(X_i + X_j) = Var(X_i) + Var(X_j)$. Also using $Var(aX + b) = a^2 Var(X)$ as shown in solution to 4a above, we get:

$$
\begin{aligned}
Var(\sqrt{n}(\widehat{\mu}_n - \mu)) &= nVar(\widehat{\mu}_n - \mu) \\
&= nVar(\widehat{\mu}_n) \\
&= nVar(\frac{1}{n}\sum X_i) \\
&= \frac{1}{n}Var(\sum X_i) \\
&= \frac{1}{n}\sum Var(X_i) \\
&= \frac{1}{n}n\sigma^2 \\
&= \sigma^2
\end{aligned}
$$

Question 5a

The rank of matrix is equal to the number of non-zero rows in row reduced echelon form. Applying elementary row operations (inter-exchange, replacement and scalar multiplication), we get $rank(A) = 2$. Note, r_i refers to the row of the matrix starting with 1 i.e. r_1 refers to first row.

$$
A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow{r_3 = r_3 - r_2} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 0 & 1 & -1 \end{bmatrix} \xrightarrow{r_2 = r_2 - r_1} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & 1 & -1 \end{bmatrix}
$$

$$
\xrightarrow{r_2 = -r_2/2} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{bmatrix} \xrightarrow{r_3 = r_3 - r_2} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{r_1 = r_1 - 2r_2} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}
$$

Similarly, we get $rank(B) = 2$.

$$
B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow{r_3 = r_3 - r_2} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{r_2 = r_2 - r_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -2 \\ 0 & 1 & 1 \end{bmatrix}
$$

$$
\xrightarrow{r_2 = -r_2/2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{r_3 = r_3 - r_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{r_1 = r_1 - 2r_2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}
$$

Question 5b

The basis (minimal size) of column span/space of a matrix is the set of columns in original matrix with pivot positions in row reduced echelon form. We computed the row reduced echelon form above, using that:

$$Basis(A) = \{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \}$$

$$Basis(B) = \{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \}$$

Question 6a

$$\begin{aligned} Ac &= \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 0+2+4 \\ 2+4+2 \\ 3+3+1 \end{bmatrix} \\ &= \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix} \end{aligned}$$

Question 6b

$Ax = b$ can be solved by reducing to row reduced echelon form using Gaussian elimination on augmented matrix. We get $x = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$.

$$\left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 2 & 4 & 2 & -2 \\ 3 & 3 & 1 & -4 \end{array}\right] \xrightarrow{r_3=r_3-r_2} \left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 2 & 4 & 2 & -2 \\ 1 & -1 & -1 & -2 \end{array}\right]$$

$$\xrightarrow{r_2=r_2-2r_3} \left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 0 & 6 & 4 & 2 \\ 1 & -1 & -1 & -2 \end{array}\right]$$

$$\xrightarrow{r_2=r_2/6} \left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 0 & 1 & 2/3 & 1/3 \\ 1 & -1 & -1 & -2 \end{array}\right]$$

$$\xrightarrow{r_1=r_1-2r_2} \left[\begin{array}{ccc|c} 0 & 0 & 8/3 & -8/3 \\ 0 & 1 & 2/3 & 1/3 \\ 1 & -1 & -1 & -2 \end{array}\right]$$

$$\xrightarrow{r_3=r_3+r_2} \left[\begin{array}{ccc|c} 0 & 0 & 8/3 & -8/3 \\ 0 & 1 & 2/3 & 1/3 \\ 1 & 0 & -1/3 & -5/3 \end{array}\right]$$

$$\xrightarrow{r_3<->r_1} \left[\begin{array}{ccc|c} 1 & 0 & -1/3 & -5/3 \\ 0 & 1 & 2/3 & 1/3 \\ 0 & 0 & 8/3 & -8/3 \end{array}\right]$$

$$\xrightarrow{r_3=3r_3/8} \left[\begin{array}{ccc|c} 1 & 0 & -1/3 & -5/3 \\ 0 & 1 & 2/3 & 1/3 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

$$\xrightarrow{r_1=r_1+r_3/3} \left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 2/3 & 1/3 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

$$\xrightarrow{r_2=r_2-2r_3/3} \left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

Question 7a

Question 7b

Question 7c

Question 8a

$M = diag(v)$ indicates a diagonal matrix formed using the entries from vector v i.e. $m_{i,i} = v_i$. The inverse of a matrix satisfies $MM^{-1} = I$, though M is a diagonal matrix, it is not clear if M^{-1} is also a diagonal matrix. Consider, $N = M^{-1}$, then $A = MN$ can be computed as:

$$\begin{aligned}
a_{i,j} &= \sum m_{i,k} n_{k,j} \\
&= m_{i,i} n_{i,j} \\
&= m_{i,i} n_{i,j}
\end{aligned}$$

But, $A = I$, hence $a_{i,i} = 1$. This implies $m_{i,i} n_{i,i} = 1$ and hence $n_{i,i} = \frac{1}{m_{i,i}}$.

Further, $a_{i,j} = 0 : i \neq j$ so $m_{i,i} n_{i,j} = 0$. Since $m_{i,i} \neq 0$, $n_{i,j} = 0$. This implies that N, the inverse of M is also a diagonal matrix.

So, $g(v_i) = \frac{1}{v_i} = w_i$ satisfies $diag(v)^{-1} = diag(w)$

Question 8b

The norm of vectors v is defined as $||v||_p = [\sum v_i^p]^{1/p}$. In this case we are dealing second norm. This can also be represented using vector multiplication as $[v^T v]^{1/p}$.

Notice, Ax is a vector in R^n. Say $v = Ax$, then using $(AB)^T = B^T A^T$ and associativity of matrix multiplication, we get:

$$\begin{aligned}
||Ax||_2^2 &= ||v||_2^2 \\
&= v^T v \\
&= (Ax)^T Ax \\
&= x^T A^T Ax \\
&= x^T x \\
&= ||x||_2^2
\end{aligned}$$

Question 8c

Using $(AB)^T = B^T A^T$, we get $(B^{-1})^T = B^{-1}$ hence B^{-1} is also symmetric.

$$\begin{aligned}
(B^{-1})^T &= (B^{-1})^T B B^{-1} \\
&= (B^{-1})^T B^T B^{-1} \\
&= (B B^{-1})^T B^{-1} \\
&= B^{-1}
\end{aligned}$$

Question 8d

By definition of eigen vectors, we have $Cy = \lambda y$ i.e. y is the eigen vector of C such that Cy is equal to a scalar multiple of y. λ is the eigen value. Substituting this in the PSD definition $x^T Cx \geq 0$ and using $v^T v = ||v||_2^2$ from 8b, we get:

$$\begin{aligned} x^T Cx &= y^T \lambda y \\ &= \lambda y^T y \\ &= \lambda ||y||_2^2 \end{aligned}$$

So we have $\lambda ||y||_2^2 \geq 0$ and since $||y||_2^2 > 0$, $\lambda \geq 0$. Hence eigen value λ is non-negative.