AML 5201

Advanced Applications of Probability & Statistics

Even Sem. 2022

Problem Set-2

Consider the following data matrix:

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

1. The projection of a sample vector $x^{(i)}$ along a direction specified by a vector v is $\left(v^{\mathrm{T}}x^{(i)}\right)/\|v\|$. Calculate the projection of the samples along the direction specified by the following vectors:

$$v = \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}-1\\-1\\-1\end{bmatrix}.$$

2. Note that the projection of a sample can also be written as $\left(\left(x^{(i)}\right)^{\mathrm{T}} v\right)/\|v\|$. If the vector v has unit magnitude, that is, $\|v\| = 1$, then the projection is simply a dot product $\left(x^{(i)}\right)^{\mathrm{T}} v$. So, we assume that the vector v has unit magnitude or convert it into a vector with unit magnitude; for example, go from $v = \begin{bmatrix}1\\1\\1\end{bmatrix}$ to $v = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}$ before calculating the projections as $\left(x^{(i)}\right)^{\mathrm{T}} v$. Then, we can write down the projections of the four samples in a vector form as follows:

$$\begin{bmatrix}\left(x^{(1)}\right)^{\mathrm{T}} v\\ \left(x^{(2)}\right)^{\mathrm{T}} v\\ \left(x^{(3)}\right)^{\mathrm{T}} v\\ \left(x^{(4)}\right)^{\mathrm{T}} v\end{bmatrix}.$$

   The quantity above is the same as (choose one): Xv, $X^{\mathrm{T}}v$, $v^{\mathrm{T}}X$, $v^{\mathrm{T}}X^{\mathrm{T}}$.

3. Calculate the mean sample from the data matrix. That is,

$$\mu = \frac{x^{(1)} + x^{(2)} + x^{(3)} + x^{(4)}}{4}.$$

4. The mean sample μ can also be calculated as (choose one):

$$\frac{1}{n}X\mathbf{1},\ \frac{1}{n}X^{\mathrm{T}}\mathbf{1},\ \frac{1}{n}\mathbf{1}^{\mathrm{T}}X,\ \frac{1}{n}\mathbf{1}^{\mathrm{T}}X^{\mathrm{T}},$$

   where $\mathbf{1}$ is the vector full of ones. In order to see this, note that:

$$\mu = \frac{x^{(1)}\times 1 + x^{(2)}\times 1 + x^{(3)}\times 1 + x^{(4)}\times 1}{4}$$

   and relate to a dot product of two things.

5. Calculate the mean of the projected samples where the projection is on to the direction of the vector $v = \begin{bmatrix}1\\1\\1\end{bmatrix}$.

6. Calculate the projection of the mean sample μ on to the direction of the vector $v = \begin{bmatrix}1\\1\\1\end{bmatrix}$.

   Compare the answer to that of the previous question. What is your conclusion?

7. What do we conclude from the following?

$$\frac{1}{4}\left(v^{\mathrm{T}}x^{(1)} + v^{\mathrm{T}}x^{(2)} + v^{\mathrm{T}}x^{(3)} + v^{\mathrm{T}}x^{(4)}\right) = v^{\mathrm{T}}\frac{\left(x^{(1)} + x^{(2)} + x^{(3)} + x^{(4)}\right)}{4} = v^{\mathrm{T}}\mu.$$

8. Calculate the variance of the projected samples where the projection is on to the direction of the vector $v = \begin{bmatrix}1\\1\\1\end{bmatrix}$.

9. What does the following quantity represent?

$$\frac{1}{n}\sum_{i=1}^{n}\left(v^{\mathrm{T}}x^{(i)} - v^{\mathrm{T}}\mu\right)^2.$$

10. We expand the quantity from the previous step as follows (fill in the blanks):

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(v^{\mathrm{T}}x^{(i)} - v^{\mathrm{T}}\mu\right)^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(v^{\mathrm{T}}x^{(i)} - v^{\mathrm{T}}\mu\right)\times\left(v^{\mathrm{T}}x^{(i)} - v^{\mathrm{T}}\mu\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(v^{\mathrm{T}}x^{(i)} - v^{\mathrm{T}}\mu\right)\times\left(\boxed{?}^{\mathrm{T}}v - \boxed{?}^{\mathrm{T}}v\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\boxed{?}\left(x^{(i)} - \mu\right)\times\left(\left(x^{(i)}\right)^{\mathrm{T}} - \mu^{\mathrm{T}}\right)\boxed{?}\right]\\
&= v^{\mathrm{T}}\left[\frac{1}{n}\sum_{i=1}^{n}\left(x^{(i)} - \mu\right)\left(\boxed{?} - \boxed{?}\right)^{\mathrm{T}}\right]v
\end{aligned}$$

11. We focus on the middle term that we derived at the end of the previous question. Fill in the blanks in the following (where we use the fact that $(a-b)^{\mathrm{T}} = a^{\mathrm{T}} - b^{\mathrm{T}}$):

$$\frac{1}{n}\sum_{i=1}^{n}\left(x^{(i)}-\mu\right)\left(x^{(i)}-\mu\right)^{\mathrm{T}} = \frac{1}{n}\left[\boxed{?}-\mu \quad \boxed{?}-\mu \quad \ldots \quad \boxed{?}-\mu\right] \times \begin{bmatrix} \left(x^{(?)}-\boxed{?}\right)^{\mathrm{T}} \\ \left(x^{(?)}-\boxed{?}\right)^{\mathrm{T}} \\ \vdots \\ \left(x^{(n)}-\boxed{?}\right)^{\mathrm{T}} \end{bmatrix}$$

$$= \frac{1}{n}\left(\begin{bmatrix} x^{(1)} & x^{(2)} & \ldots & x^{(n)} \end{bmatrix} - \mu\begin{bmatrix} \boxed{?} & \boxed{?} & \ldots & \boxed{?} \end{bmatrix}\right)\left(\begin{bmatrix} \boxed{?}^{\mathrm{T}} \\ \boxed{?}^{\mathrm{T}} \\ \vdots \\ \boxed{?}^{\mathrm{T}} \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}\boxed{?}^{\mathrm{T}}\right)$$

$$= \left(X^{\mathrm{T}} - \mu\mathbf{1}^{\mathrm{T}}\right)\left(\boxed{?} - \mathbf{1}\mu^{\mathrm{T}}\right).$$

Now we use the following facts:

- $\mu = \frac{1}{n}X^{\mathrm{T}}\mathbf{1}$,
- I represents the identity matrix with $IX = I$ and $XI = I$,
- $(ab)^{\mathrm{T}} = b^{\mathrm{T}}a^{\mathrm{T}}$,

to get

$$\frac{1}{n}\sum_{i=1}^{n}\left(x^{(i)}-\mu\right)\left(x^{(i)}-\mu\right)^{\mathrm{T}} = \frac{1}{n}\left(X^{\mathrm{T}} - \left(\frac{1}{n}X^{\mathrm{T}}\mathbf{1}\right)\mathbf{1}^{\mathrm{T}}\right)\left(X - \mathbf{1}\left(\frac{1}{n}X^{\mathrm{T}}\mathbf{1}\right)^{\mathrm{T}}\right)$$

$$= \frac{1}{n}\left(X^{\mathrm{T}} - \left(\frac{1}{n}X^{\mathrm{T}}\mathbf{1}\right)\mathbf{1}^{\mathrm{T}}\right)\left(X - \frac{1}{n}\mathbf{1}\boxed{?}^{\mathrm{T}}X\right)$$

$$= \frac{1}{n} \times \boxed{?}\underbrace{\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)}_{\text{See next question}}\boxed{?}.$$

12. Complete the steps below (note how the order of multiplication is maintained):

$$\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right) = I - I \times \left(\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right) - \left(\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right) \times I + \left(\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\left(\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)$$

$$= I - \frac{2}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}} + \frac{1}{n^2}\mathbf{1}\left(\underbrace{\mathbf{1}^{\mathrm{T}}\mathbf{1}}_{=?}\right)\mathbf{1}^{\mathrm{T}}$$

$$= I - \frac{1}{n}\boxed{?}\boxed{?}^{\mathrm{T}}.$$

13. Now use the results from (9), (10), (11) and (12) to show that the variance of the projected samples where the projection is on to the direction of a vector v is:

$$\frac{1}{n}\sum_{i=1}^{n}\left(v^{\mathrm{T}}x^{(i)} - v^{\mathrm{T}}\mu\right)^2 = v^{\mathrm{T}}\left[\frac{1}{n}\sum_{i=1}^{n}\left(x^{(i)}-\mu\right)\times\left(x^{(i)}-\mu\right)^{\mathrm{T}}\right]v = v^{\mathrm{T}}\left(\underbrace{\frac{1}{n}X^{\mathrm{T}}\boxed{?}X}_{\text{Covariance matrix}}\right)\text{v.}$$

Now principal component analysis (PCA) is about finding the vector v that maximizes the variance of the projected samples given by the last term above. We will see that the vector v will turn out be the so called *eigenvector* of the covariance matrix.