Statistical Methods in AI

Assignment 2

Megha Bose
2019111021

---

1. A. Among Eigen Value Decomposition and Singular Value Decomposition, which one is more generalizable to matrices and why?

   B. Show the method and find the Singular Value Decomposition of the following matrix: $\begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix}$

Ans: A. Single Value Decomposition (**SVD**) ($A = U\Sigma V^T$) is more generalizable than Eigen Value Decomposition ($A = PDP^{-1}$) as:

1. SVD always exists for any sort of rectangular matrix, whereas the Eigen Value Decomposition can only exist for invertible square matrices.
2. The U and V matrices we get in SVD are orthogonal and hence represent simple rotations whereas the matrices P and P^{-1} we get from eigen value decomposition need not be orthogonal.
3. In Eigen value decomposition, the entries in the diagonal matrix need to be non negative real numbers whereas there is no such restriction in SVD.

B. The steps to find Singular Value Decomposition of the matrix go as follows:

1. We first need to find the eigenvalues of $X = AA^T$ and then find matrix U by solving eigenvectors of X as they form the columns of U.
2. The singular values are the diagonal entries (σ_i) of the Σ matrix and are arranged in descending order.
3. Since U and V are orthogonal in ($A = U\Sigma V^T$), we can get columns of V (v_i) from A, singular values and columns of U (u_i) $U^TA = \Sigma V^T \implies (U^TA)^T = (\Sigma V^T)^T \implies A^TU = V\Sigma^T$ as $\frac{1}{\sigma_i} * A^T * u_i$.

Finding SVD of $A = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix}$, finding U, Σ, V s.t. $A = U\Sigma V^T$

$$A^T = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

Calculating AA^T,

$$AA^T = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 16+64 & 44+56 & 56-16 \\ 44+56 & 121+49 & 154-14 \\ 56-16 & 154-14 & 196+4 \end{bmatrix}$$

$$= \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

Now, calculating eigen values of AA^T,

$$\begin{bmatrix} 80-\lambda & 100 & 40 \\ 100 & 170-\lambda & 140 \\ 40 & 140 & 200-\lambda \end{bmatrix}$$

$$\begin{vmatrix} 80-\lambda & 100 & 40 \\ 100 & 170-\lambda & 140 \\ 40 & 140 & 200-\lambda \end{vmatrix} = 0$$

$$\Rightarrow \begin{vmatrix} 80-\lambda & 0 & 40 \\ 100 & -\lambda+170 - \dfrac{10000}{\lambda-80} & 140 \\ 40 & 140+\dfrac{4000}{\lambda-80} & 200-\lambda \end{vmatrix} = 0$$

$$\left(C_2 \leftarrow C_2 - \frac{100}{80-\lambda} C_1\right)$$

$$\Rightarrow \begin{vmatrix} 80-\lambda & 0 & 0 \\ 100 & -\lambda+170 + \dfrac{10000}{\lambda-80} & 140+\dfrac{4000}{\lambda-80} \\ 40 & 140+\dfrac{4000}{\lambda-80} & -\lambda+200 + \dfrac{1600}{\lambda-80} \end{vmatrix} = 0$$

$$\left(C_3 \leftarrow C_3 - \frac{40}{80-\lambda} C_1\right)$$

$$\Rightarrow (80-\lambda) \begin{vmatrix} -\lambda+170+\dfrac{10000}{\lambda-80} & 140+\dfrac{4000}{\lambda-80} \\ 140+\dfrac{4000}{\lambda-80} & -\lambda+200+\dfrac{1600}{\lambda-80} \end{vmatrix}$$

$$+ 0 + 0 = 0$$

$$\Rightarrow (80-\lambda)\left(\left(-\lambda+170+\frac{10000}{\lambda-80}\right)\left(-\lambda+200+\frac{1600}{\lambda-80}\right) - \left(140+\frac{4000}{\lambda-80}\right)\cdot\left(140+\frac{4000}{\lambda-80}\right)\right)$$

$$= 0$$

$$\Rightarrow (80-\lambda)$$

$$\left(\left(\lambda^2 - 200\lambda + \frac{1600\lambda}{\lambda-80} - 170\lambda + 34000 + \frac{272000}{\lambda-80} + \frac{2000000}{\lambda-80} - \frac{1000\lambda}{\lambda-80} + \cancel{\frac{16000000}{(\lambda-80)^2}}\right)\right.$$

$$\left. - \left(19600 + \cancel{\frac{16000000}{(\lambda-80)^2}} + \frac{112000}{\lambda-80}\right)\right) = 0$$

$$\Rightarrow (80-\lambda)\left(\frac{\lambda^3}{\lambda-80} + \frac{32400\lambda}{\lambda-80} - \frac{450\lambda^2}{\lambda-80}\right) = 0$$

$$\Rightarrow -(\lambda^3 - 450\lambda^2 + 32400\lambda) = 0$$

$$\Rightarrow -\lambda(\lambda^2 - 450\lambda + 32400) = 0$$

$$\Rightarrow -\lambda(\lambda-360)(\lambda-90) = 0$$

$\therefore \lambda = 0, 360, 90 \longleftarrow$ Eigen Values

4

(1) $\lambda = 360$

$$\begin{bmatrix} 80-360 & 100 & 40 \\ 100 & 70-360 & 140 \\ 40 & 140 & 200-360 \end{bmatrix} = \begin{bmatrix} -280 & 100 & 40 \\ 100 & -190 & 140 \\ 40 & 140 & -160 \end{bmatrix}$$

Let eigen vector be $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$-280x_1 + 100x_2 + 40x_3 = 0$$
$$100x_1 - 190x_2 + 140x_3 = 0$$
$$40x_1 + 140x_2 - 160x_3 = 0$$

$$-28x_1 + 10x_2 + 4x_3 = 0 \quad \text{— (1)}$$
$$10x_1 - 19x_2 + 14x_3 = 0 \quad \text{— (2)}$$
$$4x_1 + 14x_2 - 16x_3 = 0 \quad \text{— (3)}$$

Using Gaussian elimination,

$$\left[\begin{array}{ccc|c} -28 & 10 & 4 & 0 \\ 10 & -19 & 14 & 0 \\ 4 & 14 & -16 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & -5/14 & -1/7 & 0 \\ 10 & -17 & 14 & 0 \\ 4 & 14 & -16 & 0 \end{array}\right]$$

$(R_1/-28 \to R_1)$

$$\to \left[\begin{array}{ccc|c} 1 & -5/14 & -1/7 & 0 \\ 0 & -108/7 & 108/7 & 0 \\ 0 & 108/7 & -108/7 & 0 \end{array}\right] \quad \begin{pmatrix} R_2 - 10R_1 \to R_2 \\ R_3 - 4R_1 \to R_3 \end{pmatrix}$$

$$\to \left[\begin{array}{ccc|c} 1 & -5/14 & -1/7 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 108/7 & -108/7 & 0 \end{array}\right] \quad \left(R_2/-108/7 \to R_2\right)$$

$$\to \left[\begin{array}{ccc|c} 1 & 0 & -1/2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \begin{pmatrix} R_1 + 5/14 R_2 \to R_1 \\ R_3 - 108/7 R_2 \to R_3 \end{pmatrix} = \boxed{\therefore \text{ eigen vect } \begin{bmatrix} 1/2 \\ 1 \\ 1 \end{bmatrix}}$$

(2) $\lambda = 90$

$$\begin{bmatrix} 80-90 & 100 & 40 \\ 100 & 170-\lambda & 140 \\ 40 & 140 & 200-\lambda \end{bmatrix} = \begin{bmatrix} -10 & 100 & 40 \\ 100 & 80 & 140 \\ 40 & 140 & 110 \end{bmatrix}$$

Let eigen vector be $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$-10x_1 + 100x_2 + 40x_3 = 0$$
$$100x_1 + 80x_2 + 140x_3 = 0$$
$$40x_1 + 140x_2 + 110x_3 = 0$$

$$-x_1 + 10x_2 + 4x_3 = 0$$
$$10x_1 + 8x_2 + 14x_3 = 0$$
$$4x_1 + 14x_2 + 11x_3 = 0$$

Using Gaussian elimination,

$$\left[\begin{array}{ccc|c} -1 & 10 & 4 & 0 \\ 10 & 8 & 14 & 0 \\ 4 & 14 & 11 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & -10 & -4 & 0 \\ 10 & 8 & 14 & 0 \\ 4 & 14 & 11 & 0 \end{array}\right]$$

$(R_1/_{-1} \to R_1)$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & -10 & -4 & 0 \\ 0 & 108 & 54 & 0 \\ 0 & 54 & 27 & 0 \end{array}\right] \quad \begin{pmatrix} R_2 - 10R_1 \to R_2 \\ R_3 - 4R_1 \to R_3 \end{pmatrix}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & -10 & -4 & 0 \\ 0 & 1 & 1/2 & 0 \\ 0 & 54 & 27 & 0 \end{array}\right] \quad (R_2/108 \to R_2)$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \begin{pmatrix} R_1 + 10R_2 \to R_1 \\ R_3 - 54R_2 \to R_3 \end{pmatrix}$$

$$\therefore \boxed{\text{Eigen vector} = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix}}$$

(3) $\lambda = 0$

$$\begin{bmatrix} 80-0 & 100 & 40 \\ 100 & 170-0 & 140 \\ 40 & 140 & 200-0 \end{bmatrix} \therefore \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

Let eigen vector be $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$80x_1 + 100x_2 + 40x_3 = 0$$
$$100x_1 + 170x_2 + 140x_3 = 0$$
$$40x_1 + 140x_2 + 200x_3 = 0$$

$$\hookrightarrow \quad 8x_1 + 10x_2 + 4x_3 = 0$$
$$10x_1 + 17x_2 + 14x_3 = 0$$
$$4x_1 + 14x_2 + 20x_3 = 0$$

Using Gaussian Elimination,

$$\left[\begin{array}{ccc|c} 8 & 10 & 4 & 0 \\ 10 & 17 & 14 & 0 \\ 4 & 14 & 20 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 5/4 & 1/2 & 0 \\ 10 & 17 & 4 & 0 \\ 4 & 14 & 0 & 0 \end{array}\right]$$

$(R_1/8 \to R_1)$

$$\to \left(\begin{array}{ccc|c} 1 & 5/4 & 1/2 & 0 \\ 0 & 9/2 & 9 & 0 \\ 0 & 9 & 18 & 0 \end{array}\right) \quad \begin{pmatrix} R_2 - 10R_1 \to R_2 \\ R_3 - 4R_1 \to R_3 \end{pmatrix}$$

$$\to \left[\begin{array}{ccc|c} 1 & 5/4 & 1/2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 9 & 18 & 0 \end{array}\right] \quad (R_2 \times 2/9 \to R_2)$$

$$\to \left[\begin{array}{ccc|c} 1 & 0 & -2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \begin{pmatrix} R_1 - 5/4 R_2 \to R_1 \\ R_3 - 9R_2 \to R_3 \end{pmatrix}$$

$$\therefore \boxed{\text{Eigen vector} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}}$$

7

$\sigma_1 = \sqrt{360} = 6\sqrt{10}$

$\sigma_2 = \sqrt{90} = 3\sqrt{10}$

(sqrt of non-zero eigen values)

$\therefore$ Σ matrix

$$\boxed{\Sigma = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \\ 0 & 0 \end{bmatrix}}$$

Now, calculating U from normalised eigen vectors as columns

$$U = \begin{bmatrix} \frac{1/2}{3/2} & -1/\tfrac{3}{2} & 2/3 \\ 1/\tfrac{3}{2} & \frac{-1/2}{3/2} & -2/3 \\ 1/\tfrac{3}{2} & 1/\tfrac{3}{2} & 1/3 \end{bmatrix}$$

$$\Rightarrow \boxed{U = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}}$$

Now, we calculate columns of the V matrix, Let the V columns be v_1 & v_2 and of U be u_1 & u_2

$$v_1 = \frac{1}{\sigma_1}\begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix}^T u_1 = \frac{1}{6\sqrt{10}}\begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \cdot \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

$$= \frac{1}{6\sqrt{10}}\begin{bmatrix} 18 \\ 6 \end{bmatrix} = \begin{bmatrix} \frac{3\sqrt{10}}{10} \\ \frac{\sqrt{10}}{10} \end{bmatrix}$$

$$v_2 = \frac{1}{\sigma_2}\begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix}^T \cdot u_2$$

$$= \frac{1}{3\sqrt{10}}\begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}\begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}$$

$$= \frac{1}{3\sqrt{10}}\begin{bmatrix} 3 \\ -9 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{10}}{10} \\ -\frac{3\sqrt{10}}{10} \end{bmatrix}$$

$$\therefore \boxed{V = \begin{bmatrix} \frac{3\sqrt{10}}{10} & \frac{\sqrt{10}}{10} \\ \frac{\sqrt{10}}{10} & -\frac{3\sqrt{10}}{10} \end{bmatrix}}$$

Ans: $$\boxed{\Sigma = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \\ 0 & 0 \end{bmatrix}, \quad U = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}, \quad V = \begin{bmatrix} \frac{3\sqrt{10}}{10} & \frac{\sqrt{10}}{10} \\ \frac{\sqrt{10}}{10} & \frac{-3\sqrt{10}}{10} \end{bmatrix}}$$

2. A. Suppose you want to apply PCA to your data X which is in 2D and you decompose X as UDV^T. Then, which of the following are correct:
(a) PCA can be useful if all elements of D are equal
(b) PCA can be useful if all elements of D are not equal
(c) D is not full-rank if all points in X lie on a straight line
(d) V is not full-rank if all points in X lie on a straight line
(e) D is not full-rank if all points in X lie on a circle

B. True/False
PCA will project the data points(multi-class) on a line which preserve information useful for data classification.

Ans: A.

- (a) is wrong as in this case, D would contain only zeroes.
- **(b) is correct** as here D will not have all zeroes and PCA can be applied here.
- **(c) is correct** as in this case, maximum variance is achieved in the direction of the straight line on which the points lie which in turn implies that there can be at-most one non zero singular value. Since our data is in 2D, the matrix D is clearly not full rank.
- (d) There is no such condition on V.
- (e) If the points lie on a circle, the variance is spread along all the directions and hence D can be full-rank in this case.

B. The statement is **false** because PCA is used in dimensionality reduction by projecting the points onto lower dimensions while preserving but that doesn't necessarily need to be a line. It might be any dimension lower than that of the original data. For example, for 3D data, PCA might project the data points onto a plane and not a line.

3. The aim of this question is to understand Bayes Theorem. One very useful resource is Bayes Theorem video by 3b1b.

A. What is the difference between prior and posterior probabilities?

B. Let's say that you are at work one day and have just finished lunch. You suddenly feel horrible and find yourself lying down. Maybe it is because one of your friend was recently sick with flu. You have a headache and sore throat, and you know that people with the flu have the same symptoms roughly 90% of the time. In other words, 90% of people with the flu have the same symptoms you currently have. Wanting to gain a little more information you roll over, grab your phone and search Google. You find a reputable article that says that only 5% of the population will get the flu in a given year. Or, the probability of having the flu, in general, is only 5%. You then spot one more statistic that says 20% of the population in a given year will have a headache and sore throat at any given time. What is the probability of you having a flu given you have a sore throat and headache?

Ans:
A. **Prior Probability** A prior probability is the probability that an observation will fall into a group before we collect the data. The prior is a probability distribution that represents the uncertainty before we have sampled any data and attempted to estimate it.

Posterior Probability A posterior probability is the probability of assigning observations to groups given the data. It is a conditional distribution because it conditions on the observed data.

From Bayes' theorem we relate the two:

$$P(a|x) = \frac{p(x|a)P(a)}{p(x)}$$

$$Posterior\ Probability = \frac{likelihood \,\text{x}\, Prior\ Probability}{evidence}$$

where $P(a|x)$ is probability of observing class a after we have sampled x, $p(x|a)$ is the probability distribution for getting x given the observed class is a, $P(a)$ is the probability of getting class a and $p(x)$ is the probability distribution for x.

B. Let probability of getting flu be denoted by $P(flu)$ and probability of having the symptoms be $P(symp)$. Therefore, it is given that,

$$P(flu) = 0.05,\ P(symp) = 0.2$$

$$P(symp|flu) = 0.9$$

By Bayes' theorem,

$$P(flu|symp) = \frac{P(symp|flu)P(flu)}{P(symp)} = \frac{0.9 * 0.05}{0.2} = 0.225$$

Therefore, the probability of having a flu given one has a sore throat and headache is **0.225**.

4. Write a code to perform KNN classification on Iris dataset provided. Use the stater code for loading the train, test dataset. Report the accuracy obtained on test dataset. Do not use direct inbuilt functions. Numpy or other math libraries are allowed.

> **Ans:** Code in Q4.ipynb.
>
> Accuracy obtained (k=3) = **96.667%**

5. For the sample dataset provided, write a code to perform logistic regression. Plot a decision boundary between the two classes. Sample result image is provided. Do not use direct inbuilt functions. Numpy or other math libraries are allowed.

> **Ans:** Code in Q5.ipynb.
>
> Decision boundary is as follows (learning rate = 0.01, number of epochs = 2000):