SMAI - Assignment

Aditya-Kumar Singh ID: 2021701010

Q.1) Among Eigen Value Decomposition (EVD) & Singular Value Decomposition (SVD), SVD is more generalizable due to the following reasons:

a) ~~It exist~~ Unlike EVD, SVD exist for all sorts of matrixes (Rectangular). Again EVD has some limitations i.e. though it's applicable to square matrices it doesn't guarantee the decomposition for some reasons:

i) Eigen-values computed may not exist in the underlying field.

ii) Non-equality of "Algebraic & Geometric" multiplicity.

b) Further with SVD eigenvectors obtained are "orthogonal" which itself ~~given~~ is a boon for subsequent operations, which may not be the case with EVD. Plus the eigen-values obtained are non-negative.

Now we'll discuss why SVD exists?

* Say, 'X' is an $m \times n$ matrix (i.e. $X: \mathbb{R}^n \to \mathbb{R}^m$) with rank r (where $r \leq \min(m,n)$). Now, we claim that $X^T X$ is symmetric & positive-semidefinite (which we'll show later why so?).

* Due to it's symmetricity (of $X^T X$), it is diagonalizable with an orthonormal basis of eigen-vectors. And these eigenvalues (i.e, the square-root of them), the eigen-vectors of $X^T X$ & the eigen-vectors of $X X^T$, all constitute to give the expression $X = U \Sigma V^T$ ($V^T \to$ Eigen-vectors of $X^T X$; $U \to$ eigen-vectors of $X X^T$).

* Now we'll show how we'd arrived to the above formula?

a) Symmetricity: $(X^T X)^T = X^T (X^T)^T = X^T X$

b) PSD: Claim: (Need to prove) $X^T X$ is PSD $\Leftrightarrow v^T (X^T X) v \geq 0 \;\; \forall \, v \neq 0$ ✓

$$u^T(X^TX)v = (Xv)^T Xv = (Xv)^2 \geq 0 \quad - \square$$

c) ~~Why~~ How eigen-values ≥ 0?

Now, since X^TX is diagonalizable, let it EVD ~~form~~ be VDV^T i.e $\boxed{X^TX = VDV^T} \doteq \sum_{k=1}^{n} v_k \lambda_k v_k^T$

$$\Rightarrow u^T(X^TX)u = u^T\left(\sum_{k=1}^{n} v_k \lambda_k v_k^T\right)u$$

$$= \sum_{k=1}^{n} \lambda_k (u^T v_k v_k^T u)$$

$$= \sum_{k=1}^{n} \lambda_k (v_k^T u)^2 \geq 0$$

(∵ X^TX is PSD $\Rightarrow u^T(X^TX)u \geq 0$)

Sufficient condition :

→ If All $\lambda_k \geq 0$, the above expression is true.

But we want stricter conditions:

Consider 'v_i' as eigen-value of X^TX, then

$$v_i^T(X^TX)v_i = v_i^T(\lambda_i v_i) = \lambda_i (v_i^T v_i) \geq 0 \quad (∵ X^TX \text{ is PSD})$$

$$\Rightarrow \lambda_i \geq 0 \quad (∵ v_i^T v_i = 1 \text{ (since orthonormal)})$$

d) let $\lambda_1 = \sigma_1^2 \geq \lambda_2 = \sigma_2^2 \geq \cdots \geq \lambda_r = \sigma_r^2 \geq 0$ (since λ_i's are ≥ 0)

→ Now, why only 'r' eigen-values, because $\text{rank}(X) = \text{rank}(X^TX) = r \Rightarrow \exists$ only 'r' non-zero eigen-values while others are '0'.

→ Since $\dim(\text{Range-space}(X^TX)) = r$, then $\exists\ v_1, v_2, \ldots, v_r$ orthonormal eigen-vectors corresponding to 'r' non-zero eigen-values (i.e.. $\lambda_1, \lambda_2, \ldots, \lambda_r$) & construct $v_{r+1}, \ldots, v_n$ = orthonormal basis for eigen-value $= 0$ (basically this spans the $N(X^TX)$ & can be constructed using Gram-Schmidt orthogonalization process).

→ Define: $u_i = \frac{1}{\sigma_i} X v_i$, then '$u_i$' is the unit eigen vector of XX^T with eigen-value σ_i^2 (check).

[Now here an elementary form of SVD can be seen, i.e.

$X v_i = u_i \sigma_i$].

Construct $V_{n \times r} = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_r \\ | & | & & | \end{bmatrix}_{n \times r}$, $\Sigma_r = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ 0 & \sigma_2 \cdots & 0 \\ \vdots & & \\ 0 & \cdots & \sigma_r \end{bmatrix}_{r \times r}$

& U_r = $m \times r$ matrix with $u_i = \frac{1}{\sigma_i} X v_i$ (as i^{th} column).

Then, we have $U_r = X V_r \Sigma_r^{-1} \Rightarrow \boxed{U_r \Sigma_r = X V_r}$

$$\boxed{\Rightarrow U_r \Sigma_r V_r^T = X V_r V_r^T = X_{m \times n} I_{n \times n} = X_{m \times n}}$$

[Again, we can invoke the orthonormal vectors from null space of X^TX & in V_r & of XX^T in U_r & pack the Σ_r with 0's to obtain $\boxed{X = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T}$ = which is basically same as $\hat{U}_{m \times n} \hat{\Sigma}_{n \times n} V_{n \times n}^T$ (Truncated SVD form)

Note:

→ Again SVD can be ~~decom~~ computed using QR decomposition method to make it fast.

→ For a symmetric matrix, SVD = EVD.

B) The method for SVD is already shown above → now we'll use the motivation from above derivation to ~~gi~~ decompose the given matrix. $M = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix}$

a) let $A := M^T M = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} = \begin{bmatrix} 333 & 81 \\ 81 & 117 \end{bmatrix}$

b) for a symmetric matrix EVD = SVD.

let's compute the eigen-value of 'A' using "characteristic" equation:

$$|A-\lambda I| = |(\lambda - 333)(\lambda - 117) - 81^2| = 0$$

$$\Rightarrow \quad \lambda = \underset{\lambda_1}{360}, \underset{\lambda_2}{90}$$

$$\Rightarrow Av_1 = \lambda_1 v_1 = \lambda_1 \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix}$$

$$\Rightarrow 9(37 v_{11} + 9 v_{12}) = 360 v_{11} \quad \& \quad 9(9v_{11} + 13 v_{12}) = 360 v_{12}$$

$$\Rightarrow v_{11} = 3 v_{12} \Rightarrow v_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = v_1 \text{ (normalized)}$$

$$\Rightarrow Av_2 = \lambda_2 v_2 = \lambda_2 \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix}$$

$$\Rightarrow 9(37 v_{21} + 9 v_{22}) = 90 v_{21} \quad \& \quad 9(9v_{21} + 13 v_{22}) = 90 v_{22}$$

$$\Rightarrow v_{21} = -\frac{1}{3} v_{12} \Rightarrow v_2 = \begin{bmatrix} 1 \\ -3 \end{bmatrix} \Rightarrow \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ -3 \end{bmatrix} = v_2$$

(normalized).

→ Verify that $v_1^T v_2 = 0$ (orthonormal)

→ Define:
$$u_1 = \frac{1}{\sigma_1} M v_1 = \frac{1}{6\sqrt{10}} \cdot \frac{1}{\sqrt{10}} \begin{bmatrix} 20 \\ 40 \\ 40 \end{bmatrix}$$

$$= \frac{1}{3} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$$u_2 = \frac{1}{\sigma_2} M v_2 = \frac{1}{3\sqrt{10}} \cdot \frac{1}{\sqrt{10}} \begin{bmatrix} 20 \\ 10 \\ -20 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$$

(~~note~~ Again $u_1^T u_2 = 0$ → orthonormal)

→ Now let's check: if
$$\underbrace{\begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sqrt{360} & 0 \\ 0 & \sqrt{90} \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} | & | \\ v_1 & v_2 \\ | & | \end{bmatrix}}_{V}^T = M$$

→

$$\frac{\sqrt{90}}{3\sqrt{10}} \underbrace{\begin{bmatrix}1&2\\2&1\\2&-2\end{bmatrix}}_{U}\underbrace{\begin{bmatrix}2&0\\0&1\end{bmatrix}}_{\Sigma}\underbrace{\begin{bmatrix}3&1\\-1&+3\end{bmatrix}}_{V^T}$$

$$= \begin{bmatrix}1&2\\2&1\\2&-2\end{bmatrix}\begin{bmatrix}6&2\\-1&+3\end{bmatrix} = \begin{bmatrix}4&8\\11&7\\14&-2\end{bmatrix} = M$$

Hence $\boxed{M = \underset{3\times 2}{\begin{bmatrix}1&2\\2&1\\2&-2\end{bmatrix}}\underset{2\times 2}{\begin{bmatrix}2&0\\0&1\end{bmatrix}}\underset{2\times 2}{\begin{bmatrix}3&1\\-1&3\end{bmatrix}}}$ = Truncated SVD form = Same as full SVD.

Q2) A) Since 'X' is in 2D ⇒ 'X' data has '2' features.

& $[X^TX]_{2\times 2}$ (where ~~$X \in \mathbb{R}^{m\times 2}$~~ $\dim(X) = m \times 2$, m = no. of data points)

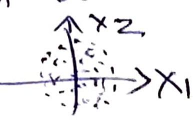
a) If all elements of 'D' are equal can be interpreted in '2'-ways:

1) $D = \begin{bmatrix}0&0\\0&0\end{bmatrix}$ ⇒ $\text{Rank}(D)=0$ ⇒ All vectors in $\mathbb{R}^2$ lies in Null-space of X^TX ⇒ $(X^TX)v = 0 \ \forall\ v \in \mathbb{R}^2$

So X^TX = zero map & as $R(X) = R(X^TX) = 0$

⇒ $R(X) = 0$ ⇒ 'X' consists of 'm' zero-vectors.

2) $D = \begin{bmatrix}\lambda_1&0\\0&\lambda_1\end{bmatrix}$ ⇒ $\text{Rank}(D)=2$ ⇒ Both eigen-vectors (orthonormal ofcourse) are important principal directions, so projecting onto one would result in huge loss of info. And here if we visualize the data, they can be seen to scattered in "spherical" manner



So, no PCA can't be useful if all elements of D are equal.

b) True, if $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, with $\lambda_1 > \lambda_2 \Rightarrow$ The eigen-vector corresponding to 'λ_1' (i.e. V_1) is the direction along which maximum variance is captured. And hence the data point can be projected onto that line.

c) ~~False~~ ~~True~~ (True), since all points in 'X' are lying on a straight line $\Rightarrow X = \begin{bmatrix} a & b \\ 2a & 2b \\ \vdots & \\ ma & nb \end{bmatrix} \Rightarrow Rank(\hat{X}) = 1 = Rank(\hat{X}^T\hat{X})$

(where $\hat{X} = X - \tilde{\mu}$ = mean-shifted

~~$\Rightarrow$ Since D^2 = Eigen-value matrix of X^TX, then.~~

$\Rightarrow$ ~~$N(X^TX)$~~ $Nullity(X^TX) = \dim(\text{Null-space}(X^TX))$

$= 2 - Rank(X^TX)$ (from rank nullity)

$= 1$

$\Rightarrow \exists$ only one (upto span) $v (\neq 0) \in \mathbb{R}^2$ s.t. $(X^TX)v = 0 \cdot v = 0$

Hence '0' is an eigen-value

$\Rightarrow D = \begin{bmatrix} \sqrt{\lambda} & 0 \\ 0 & 0 \end{bmatrix}$ where ~~$\lambda \neq 0$~~ $\lambda > 0$. (exist as Rank = 1) $\hat{X} = \begin{bmatrix} -1.5 & 0 \\ -0.5 & 0 \\ 0.5 & 0 \\ 1.5 & 0 \end{bmatrix}$

(λ is eig-value of X^TX) where $\hat{X} = X - \tilde{\mu}$ (mean shifted)

$\Rightarrow$ Hence D is not full rank.

$\Rightarrow$ And 'V_1' is the direction (or line) on which they '$\hat{X}$' are lying while 'V_2' is the $\perp$ vec to V_1.

$\rightarrow$ But if $X = \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 3 & 2 \\ 4 & 2 \end{bmatrix}$ i.e. $y = 2$ line, then $\boxed{Rank(\hat{X}) = Rank(\hat{X}^T\hat{X}) = Rank(D) = 1}$

d) False, Since Nullity = 1 & Rank = 1, one eigen-vector would be from Range-space of X^TX & another one from Null-space (since '0' is an eigen-value). And in this case

$$\text{Range-space}(X^TX) = [\text{Null-space}(X^TX)]^\perp$$

as due to spectral theorem, symmetric X^TX would yield orthogonal eigen-vectors.

Now $V = \begin{bmatrix} | & | \\ v_1 & v_2 \\ | & | \end{bmatrix}_{2\times 2}$, where both are independent $\Rightarrow$ V' is full rank.

e) False,

Since it is being said they are lying on a circle (presumably ~~there~~ $\nexists$ only two points in X which are collinear & supposedly lying on circle) then $\exists$ two such vectors which are independent $\Rightarrow$ ~~Applying~~ Rank$(X) = 2 \Rightarrow$ Rank(X^TX) $=$ Rank$(X) = 2 \Rightarrow$ Null-space$(X^TX) = \{0\}$ = zero vector

$\Rightarrow \nexists$ ~~do~~ any zero-eigen value $\Rightarrow D = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix}$

where $\lambda_1, \lambda_2 \neq 0$.

$\Rightarrow D'$ is full-rank.

B) False. Since PCA relies on SVD & SVD doesn't invoke any class labels, rather ~~do~~ we only the data-information to find suitable max-varying direction to preserve information, ~~the we won't~~ it won't be useful for data classification. Again, it may so happen that the direction of max-variability is the direction of max-separation bet^n classes which is just a mere co-incidence.

3)

A) ~~Prior~~ According to Bayes, formula:

if H = Hypothesis & E = Evidence

$$P(H|E) = \frac{P(E|H)\, P(H)}{P(E)}$$

$\rightarrow$ Here, $P(H)$ = prior (mathematically) & $P(H|E)$ = posterior (mathematically)

Now, intuitively what it means is that : (or the statement we're proposing)

* $P(H)$ = probability that hypothesis hold (e.g. in our case, according to question, the hypothesis is made about happening of flu to a given person in a given year) without considering any evidence

* $P(H|E)$ = "posterior"

= It's the Belief about the hypothesis we proposed after seeing the evidence. In our case it would be framed as follows :

"If we have "evidence" that person is having 'headache & sore throat', then what should be the likelihood ~~of catching flu~~ that that person has flu, OR in other words, what is the probability that our hypothesis is true or what is the confidence on our hypothesis.

B) Ans Given:

1) $P(ST\&H \mid Flu) = \frac{90}{100}$ ($ST\&H$ = sore throat & headache)

2) $P(Flu) = \frac{5}{100}$

3) $P(ST\&H) = 20/100$

To be found : $P(Flu \mid ST\&H)$.

Using Bayes Rule :

$$P(Flu \mid ST\&H) = \frac{P(ST\&H \mid Flu) \cdot P(Flu)}{P(ST\&H)}$$

$$= \frac{90/100 \cdot 5/100}{20/100} = \frac{90 \cdot 5}{20 \cdot 100}$$

$$= \frac{45}{200} = \boxed{22.5\ \%} \text{ or } \boxed{0.225}$$