Name: Tsering Jorgais
SR- No: 19648

MLSP - Assignment 1

Solution 1 - Induction in PCA

→ We know that, In order to maximize the variance of 1-dim Projection $y = w^T x$ of D-dimensional data $x$, we need to find $w = v_1$, Such that $v_1$ is eigen vector corresponding to highest eigen value of the Covariance matrix

$$S = \frac{1}{N} \sum_{n=1}^{N} (x - \bar{x})(x - \bar{x})^T.$$

→ Lets assume that the variance of M-dimensional projectional $y_M = w_M^T x$ is maximized by $W = [v_1, v_2, \dots, v_m]$ where $v_1, v_2, \dots, v_m$ are orthogonal eigen vector of $S$ corresponding to the M - largest eigen values.

→ To maximized the variance of M+1 dimension projection $y_{M+1} = W_{M+1}^T x$ we need to prove that $W_{M+1} = [W_M, v_{M+1}]$ where $v_{M+1}$ is eigen vector of S corresponding to ~~highest~~ $(M+1)^{th}$ largest eigen value.

The variance of Projected data is

$$S_y = \frac{1}{N} \sum_{n=1}^{N} (y_{M+1} - \bar{y})^2$$

$$\frac{1}{N} \sum_{n=1}^{N} (W_{M+1}^T x_n - W_{M+1}^T \bar{x})^2 \qquad \{ y_{M+1} = W_{M+1}^T x_n \}$$

$$\frac{1}{N} \sum_{n=1}^{N} W_{M+1}^T (x_n - \bar{x})(x_n - \bar{x})^T W_{M+1}$$

$$\frac{1}{N} W_{M+1}^T \sum_{n=1}^{N} (x_n - \bar{x})(x_n - \bar{x})^T W_{M+1}$$

$$W_{M+1}^T \underbrace{\frac{1}{N} \sum_{n=1}^{N} (x_n - \bar{x})(x_n - \bar{x})^T}_{S} W_{M+1}$$

$$W_{M+1} S W_{M+1}$$

we Can write it as.

$$\sum_{i=1}^{M+1} U_i^T S U_i$$ [ $U_i \to$ eigen vector, corresponding to eigenvalue $\lambda_i$ ]

We have to maximize the

$$\max_{U_i} \sum_{i=1}^{M+1} U_i^T S U_i$$

$$\underbrace{\sum_{i=1}^{M} U_i^T S U_i} + U_{m+1}^T S U_{m+1} \quad \text{——— (1)}$$

We already know that this term is maximized by M - largest Eigen vectors corresponding to M-largest eigenvalues, therefore we need to maximize the Second term. $U_{m+1}^T S U_{m+1}$ Such that

$U_{m+1}^T U_{m+1} = 1$ and $U_{m+1}^T U_i = 0 \quad \forall \; 1 \le i \le m$

→ Solving this Constraint Optimization Problem

Writing the Lagrangian of the Problem.

$$\mathcal{L}(\lambda_1, \lambda_2 \dots \lambda_{m+1}) = U_{m+1}^T S U_{m+1} - \lambda_{m+1}(U_{m+1}^T U_{m+1} - 1)$$

$$\nabla \mathcal{L}(\lambda_1, \lambda_2 \dots \lambda_{m+1}) = 2 S U_{m+1} - \lambda_{m+1} U_{m+1} = 0$$

$$= S U_{m+1} = \lambda_{m+1} U_{m+1}$$

from this we can say that $(M+1)^{th}$ term of $W_{m+1}$ is the eigen vector, corresponding to $(m+1)^{th}$ largest eigenvalue of S and it is orthogonal to all the eigenvector $U_i$ $\forall \; 1 \le i \le M$.

and projection vector $W_{m+1}$ $[U_1, U_2 \dots U_{m+1}]$ is first ~~first~~ $(m+1)$ ~~largest~~ eigen vector corresponding to first $(M+1)$ largest eigen values.

Solution 2 to To prove $\frac{\delta}{\delta A} Tr(AB) = 2B - diag(B)$.

$$\frac{\delta Tr(AB)}{\delta A} = \begin{bmatrix} \frac{\delta Tr(AB)}{\delta A_{11}} & \frac{\delta Tr(AB)}{\delta A_{12}} & \cdots & \frac{\delta Tr(AB)}{\delta A_{1n}} \\ \vdots & & & \\ \frac{\delta Tr(AB)}{\delta A_{n1}} & \cdots & \cdots & \frac{\delta Tr(AB)}{\delta A_{nn}} \end{bmatrix} \quad -(1)$$

$Tr(AB)$ can be written as.

$$\sum_{i=1}^{n} \sum_{j \neq i}^{n} a_{ij} b_{ji} + \sum_{i=1}^{n} a_{ij} b_{ji}$$

We can break the first term into two parts.

$a_{ij} b_{ji} + a_{ji} b_{ij}$ for a specific $(i, j)$ and they will be same due to symmetric. therefore we can say that:

$$\frac{\delta Tr(AB)}{\delta A_{ij}} = 2B_{ji} \quad \text{for non diagonal}$$

$$\frac{\delta Tr(AB)}{\delta A_{ij}} = B_{ji} \quad \text{for diagonal term}$$

from (1)

$$\frac{\delta Tr(AB)}{\delta A} = \begin{bmatrix} B_{11} & 2B_{12} & 2B_{13} & \cdots & 2B_{1n} \\ \vdots & & & & \\ 2B_{n1} & \cdots & & & B_{nn} \end{bmatrix}$$

Comment:

$\{[B_{ij} = B_{ji}]\}$.

we can write

$$\begin{bmatrix} 2B_{11}-B_{11} & 2B_{12} & 2B_{13} & \cdots & 2B_{1n} \\ \vdots & \vdots & & & \vdots \\ 2B_{n1} & \cdots & & & 2B_{nn}-B_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} 2B_{11} & 2B_{12} & \cdots & 2B_{1n} \\ \vdots & \vdots & & \vdots \\ 2B_{n1} & \cdots & & 2B_{nn} \end{bmatrix} - \begin{bmatrix} B_{11} & 0 & 0 & \cdots & 0 \\ 0 & B_{22} & & & 0 \\ 0 & & \ddots & & 0 \\ 0 & & & & \vdots \\ 0 & \cdots & & & B_{nn} \end{bmatrix}$$

$$\Rightarrow \quad 2B - \operatorname{diag}(B)$$

Solution a. To prove $\dfrac{\delta \log(|A|)}{\delta A} = 2A^{-1} - \operatorname{diag}(A^{-1})$

$$\rightarrow \frac{\delta \log(|A|)}{\delta A} = \frac{1}{|A|} \frac{\delta \det(A)}{\delta A}$$

$$\frac{\delta \det(A)}{\delta A} = \begin{bmatrix} \frac{\delta \det(A)}{\delta A_{11}} & \frac{\delta \det(A)}{\delta A_{12}} & \cdots & \frac{\delta \det(A)}{\delta A_{1n}} \\ \vdots & & & \\ \frac{\delta \det(A)}{\delta A_{n1}} & \cdots & & \frac{\delta \det(A)}{\delta A_{nn}} \end{bmatrix}$$

Solution 3. Covariance ~~of~~ of the whitenned ~~input~~ output.

$$S_T^y = \frac{1}{N}\sum_{n=1}^{N} y_n y_n^T \quad \text{--- (1)}$$

To show ~~y~~ $S_T^y = I$ where $I$ is identity matrix

We know that $y_n = \Lambda^{-1/2} W^T (x_n - \mu)$ where $W$ is eigen vector & $\Lambda$ is eigen value diagonal Matrix.

Substituting $y_n$ value in (1).

$$S_T^y = \frac{1}{N}\sum_{n=1}^{N} \Lambda^{-1/2} W^T (x_n - \mu) \cdot \left(\Lambda^{-1/2} W^T (x_n - \mu)\right)^T$$

$$= \frac{1}{N}\sum_{n=1}^{N} \Lambda^{-1/2} W^T (x_n - \mu)(x_n - \mu)^T W \left(\Lambda^{-1/2}\right)^T$$

$$= \Lambda^{-1/2} W^T \underbrace{\frac{1}{N}\sum_{n=1}^{N} (x_n - \mu)(x_n - \mu)^T}_{S} W \left(\Lambda^{-1/2}\right) \quad \left\{\Lambda^{-1/2} \text{ is diag}\right. \quad \text{--- (2)}$$

Since $S$ is the Covariance Matrix & it is Symmetric we can write it in the form of eigen decomposition

$$S = W \Lambda W^T \quad \& \left\{ W^T W = I \right\}$$

by orthogonality of eigen vectors

therefore we ~~can~~ can write (2) as.

$$\Lambda^{-1/2} \underbrace{W^T W}_{I} \Lambda \underbrace{W^T W}_{I} \Lambda^{-1/2}$$

$$\Lambda^{-1/2} \Lambda \Lambda^{-1/2} = I$$

hence Proved.

Solution 6. We first define $S_T = S_B + S_W$.

where $S_T$ is Total Variance. which is $I$ in LDA

$S_B$ is B/w Class variance

$S_W$ is within Class variance

The first LDA Projection vector $w$ is eigen vector of $S_W^{-1} S_B$ corresponding to largest eigen value and $S_W^{-1} S_B$ is positive definite matrix.

Assume $S_W$ is invertable matrix

$$S_W^{-1} S_B U_i = \lambda_i U_i \quad \text{---} \ (2)$$ $U_i \rightarrow$ eigen vector $(w = U_i)$

from (1) we can write $S_B = S_T - S_W$.

$S_B = I - S_W$.

Substituting it in (2)

$$S_W^{-1} ( I - S_W ) U_i = \lambda_i U_i$$

$$\left( S_W^{-1} - I \right) U_i = \lambda_i U_i$$

$$S_W^{-1} U_i = \lambda_i U_i + U_i$$

$$S_W^{-1} U_i = (\lambda_i + 1) U_i$$

$$S_W U_i = \left( \frac{1}{\lambda_i + 1} \right) U_i$$

where $\left( \frac{1}{\lambda_i + 1} \right)$ is eigen ~~value~~ Value. of $S_W$.

from this we have shown that the eigen vector $U_i$ is corresponding to minimum eigen value of $S_W$ since $\lambda_i$ was magnitude of maximum for $S_W^{-1} S_B$.

hence proved.