KAWIN M

E9 205 - Machine Learning for Signal Processing

Homework 1

① To Prove: The variance of M+1 dimensional projection $y_{M+1} = W_{M+1}^T x$ is maximised by choosing $W_{M+1} = [W_M \; u_{M+1}]$, where $W_M = [u_1 \, u_2 \ldots u_M]$ and $u_1, \ldots u_M, u_{M+1}$ are the orthonormal eigen vectors of S corresponding to the M+1 largest eigen values.

Proof by Induction:

Base case: [When M=1]

The variance of $y_1 = u_1^T x$ is maximised by choosing u_1, the eigen vector of S corresponding to the largest eigen value.

Proved in the lecture.

Induction Hypothesis:

Let us assume that the variance of M dimensional projection $y_M = W_M^T x$ is maximised by $W_M = [u_1 \, u_2 \ldots u_M]$ where $u_1 \ldots u_M$ are the orthonormal eigen vectors of S corresponding to the M largest eigen values $\lambda_1, \ldots \lambda_M \rightarrow (1)$

To prove, for M+1

To find a vector u_{M+1} such that the variance given by $u_{M+1}^T S_x u_{M+1}$ is maximised,

where, S_x = sample covariance matrix of the data

$$\therefore \max_{u_{M+1}} u_{M+1}^T S_x u_{M+1} \longrightarrow (2)$$

such that

1) u_{M+1} is a unit vector [$\because$ only the direction is important, not the magnitude of the vector]

$$\Rightarrow u_{M+1}^T u_{M+1} = 1 \longrightarrow (3)$$

2) u_{M+1} is orthogonal to the previously choosen M vectors $u_1 \ldots u_M$

$$\Rightarrow u_{M+1}^T u_i = 0 \longrightarrow (4) \quad \forall i = 1 \text{ to } M$$

Constrained Maximization using Lagrangian Optimization

The constraints are given using the lagrangian multipliers $\lambda_{M+1}, \lambda'_1 \ldots \lambda'_M$

Lagrangian

$$L(u_{M+1}, \lambda_{M+1}, \lambda'_1, \ldots \lambda'_M) = u_{M+1}^T S_x u_{M+1} + \lambda_{M+1}(1 - u_{M+1}^T u_{M+1}) + \sum_{i=1}^{M} \lambda'_i u_{M+1} u_i \longrightarrow (5)$$

$$\max_{u_{M+1}} L$$

Differentiating L w.r.t u_{M+1} and equating it to 0.

$$\frac{\delta L}{\delta u_{M+1}} = 0 \rightarrow (6)$$

$$\Rightarrow \frac{\delta L}{\delta u_{M+1}} = (S_x + S_x^T) u_{M+1} + \lambda_{M+1} 0 - (I + I^T) \lambda_{M+1} u_{M+1} + \sum_{i=1}^{M} \lambda'_i u_i$$

$$\left[\because \frac{\delta (x^T A x)}{\delta x} = (A + A^T) x \text{ and } \lambda u_{M+1}^T u_{M+1} = \lambda u_{M+1}^T I u_{M+1} \right]$$

$$\Rightarrow \frac{\delta L}{\delta u_{M+1}} = 2 S_x u_{M+1} - 2 \lambda_{M+1} u_{M+1} + \sum_{i=1}^{M} \lambda'_i u_i = 0 \rightarrow (7)$$

$$\left[\because S_x \text{ is symmetric} \Rightarrow S_x = S_x^T \quad \therefore S_x + S_x^T = S_x + S_x = 2 S_x \right]$$

Left multiply (7) with u_j^T, for any arbitrary $j = 1$ to M

$$\Rightarrow 0 = 2 u_j^T S_x u_{M+1} - 2 \lambda_{M+1} u_j^T u_{M+1} + u_j^T \sum_{i=1}^{M} \lambda'_i u_i$$

$\Rightarrow$ As $u_j^T \cdot u_i = 0 \quad \forall\, 1 \leq j, i \leq M+1,\ j \neq i$ and

$u_j^T \cdot u_j = 1 \quad \forall\, 1 \leq j \leq M$

$$= 2 u_j^T S_x u_{M+1} - 2 \lambda_{M+1} \cdot 0 + \lambda'_j = 0$$

$$= 2u_{M+1}^T S_x u_j + \lambda'_j = 0$$ [$\because$ S_x is symmetric $u_j^T S_x u_{M+1} = u_{M+1}^T S_x u_j$]

From the Induction Hypothesis 1

$$S_x u_j = \lambda_j u_j$$

$$\therefore = 2u_{M+1}^T \lambda_j u_j + \lambda'_j = 0$$

$$\Rightarrow 2\lambda_j u_{M+1}^T u_j + \lambda'_j = 0$$

$$\Rightarrow 2\lambda_j \cdot 0 + \lambda'_j = 0$$

$$\Rightarrow \lambda'_j = 0 \quad \text{for any } j = 1 \text{ to } M$$

$$\therefore \frac{\partial L}{\partial u_{M+1}} = 2S_x u_{M+1} - 2\lambda_{M+1} u_{M+1} = 0$$ [On substituting value of λ'_j on Eq 7]

$$\Rightarrow 2S_x u_{M+1} = 2\lambda_{M+1} u_{M+1}$$

$$\Rightarrow S_x u_{M+1} = \lambda_{M+1} u_{M+1}$$

$\therefore$ u_{M+1} is an eigen vector of S_x corresponding to eigen value λ_{M+1}

We have to find u_{M+1} such that

$$\max_{u_{M+1}} u_{M+1}^T S_x u_{M+1} = \max_{u_{M+1}} u_{M+1}^T \lambda_{M+1} u_{M+1}$$

$$= \max_{u_{M+1}} \lambda_{M+1} u_{M+1}^T u_{M+1}$$

$$= \max_{u_{M+1}} \lambda_{M+1}$$

$\therefore$ Find the eigen vector u_{M+1} such that it has the largest eigen value among those not yet choosen [i.e the $M+1$ th largest eigen value of S_x]

Hence proved

With this proof, we have PCA solution for any $M \leq D$.

Question (2) To prove

i) $\frac{\partial}{\partial A} \log(|A|) = 2A^{-1} - \text{diag}(A^{-1})$ where, A is a square symmetric matrix

Proof:

$$\frac{\partial \log(|A|)}{\partial A} = \frac{1}{|A|} \frac{\partial |A|}{\partial A}$$

Claim:

$$\frac{\partial |A|}{\partial A_{ij}} = \begin{cases} adj_{ii} & \text{if } i=j \\ 2adj_{ij} & \text{if } i \neq j \end{cases}$$

where, A_{ij} = entry of ith row, jth column of A

Proof:

Case (1) : When $i=j$

Let A be a $n \times n$ matrix.

Expanding the determinant along the ith row

$$\frac{\partial |A|}{\partial A_{ii}} = \frac{\partial}{\partial A_{ii}} (A_{i1}C_{i1} + A_{i2}C_{i2} + \ldots + A_{ii}C_{ii} + \ldots + A_{in}C_{in})$$

where C_{ij} = cofactor of A_{ij} in A

As all other entries of A are independent of A_{ii} other than itself

$$\frac{\partial |A|}{\partial A_{ii}} = 0 + \ldots + 1 \cdot C_{ii} + \ldots + 0 = C_{ii}$$

As A is a symmetric matrix,

$$\operatorname{adj} A = C$$

where,

$\operatorname{adj} A$ = Adjoint of matrix A

C = Cofactor matrix of A

$$\Rightarrow C_{ii} = \operatorname{adj}_{ii}$$

$$\therefore \frac{\partial |A|}{\partial A_{ii}} = \operatorname{adj}_{ii} \longrightarrow (1)$$

Case (2): when $i \neq j$

Expanding the determinant along the i^{th} row

$$= \frac{\partial}{\partial A_{ij}} (A_{i1} C_{i1} + A_{i2} C_{i2} + \dots + A_{ii} C_{ii} + \dots + A_{ij} C_{ij} + \dots + A_{in} C_{in})$$

Here, both A_{ij} and A_{ji} are dependent on A_{ij} and all other entries are independent

$\therefore$ Expanding all cofactors along j^{th} row

$$= \frac{\partial}{\partial A_{ij}} \Big(A_{i1} \sum_{K=2}^{n} A_{jK} C_{jK}^{A_{i1}'} + \dots + A_{ii} \sum_{K \neq i}^{n} A_{jK} C_{jK}^{A_{ii}'} + \dots +$$

$$A_{ij} \sum_{K \neq j}^{n} A_{jK} C_{jK}^{A_{ij}'} + \dots + A_{in} \sum_{K=1}^{n-1} A_{jK} C_{jn}^{A_{in}'}$$

~~wer~~ where, $C_{jK}^{A_{im}'}$ = Cofactor of A_{jK} of matrix A after removing ith row and mth column

Rewriting, by taking out A_{ji} term:

$$\frac{\partial}{\partial A_{ij}}\Big(A_{i1}\Big[A_{ji}C_{ji}^{A'_{i1}} + \sum_{\substack{K=2\\K\neq i}}^{n} A_{jK}C_{jK}^{A'_{i1}}\Big] + \dots + A_{ii}\sum_{K\neq i}^{n} A_{jK}C_{jK}^{A'_{ii}} + \dots$$

$$+ A_{ij}\Big[A_{ji}C_{ji}^{A'_{ij}} + \sum_{\substack{K=1\\K\neq j\\K\neq i}}^{n} A_{jK}C_{jK}^{A'_{ij}}\Big]$$

$$+ A_{in}\Big[A_{ji}C_{ji}^{A'_{in}} + \sum_{\substack{K=1\\K\neq i}}^{n-1} A_{jK}C_{jK}^{A'_{in}}\Big]$$

Differentiating with respecting to A_{ij}

$$= A_{ii}\Big[C_{ji}^{A'_{i1}} + 0\Big] + \dots + 0 + \dots + \Big[2A_{ij}C_{ji}^{A'_{ij}} + \sum_{\substack{K=1\\K\neq j\neq i}}^{n} A_{jK}C_{jK}^{A'_{ij}}\Big]$$

$$+ A_{in}\Big[C_{ji}^{A'_{in}} + 0\Big]$$

$[\because A_{ij} = A_{ji}]$

Rearranging

$$= A_{ii}C_{ji}^{A'_{i1}} + \dots + A_{ij}C_{ji}^{A'_{ij}} + \dots + A_{in}C_{ji}^{A'_{in}} + A_{ji}C_{ji}^{A'_{ij}} + \sum_{\substack{K=1\\K\neq j\neq i}}^{n} A_{jK}C_{jK}^{A'_{ij}}$$

$$= \Big(A_{i1}C_{ji}^{A'_{i1}} + \dots A_{ij}C_{ji}^{A'_{ij}} + \dots + A_{in}C_{ji}^{A'_{in}}\Big) + \sum_{\substack{K=1\\K\neq j}}^{n} A_{jK}C_{jK}^{A'_{ij}}$$

$$= \sum_{\substack{K=1\\K\neq i}}^{n} A_{iK}C_{ji}^{A'_{iK}} + \sum_{\substack{K=1\\K\neq j}}^{n} A_{jK}C_{jK}^{A'_{ij}}$$

Here, $\sum_{\substack{K=1\\K\neq i}}^{n} A_{iK}\cdot C_{ji}^{A'_{iK}} = \cancel{adj}_{ji}\ C_{ji}$ and $\sum_{\substack{K=1\\K\neq j}}^{n} A_{jK}C_{jK}^{A'_{ij}} = \cancel{adj_{ij}}\ C_{ij}$

$= C_{ji} + C_{ij}$

As A is symmetric

$= C_{ij} + C_{ij}$

$= 2C_{ij}$

As A is symmetric, adj $A = C$

$\therefore = 2\,\mathrm{adj}_{ij} \rightarrow (2)$

From (1) & (2)

Hence proved that

$$\frac{\partial |A|}{\partial A_{ij}} = \begin{cases} \mathrm{adj}_{ii} & \text{if } i = j \\ 2\,\mathrm{adj}_{ij} & \text{if } i \neq j \end{cases}$$

$$\therefore \frac{\partial |A|}{\partial A} = \begin{bmatrix} \frac{\partial |A|}{\partial A_{11}} & \cdots & \frac{\partial |A|}{\partial A_{1n}} \\ \vdots & & \vdots \\ \frac{\partial |A|}{\partial A_{n1}} & \cdots & \frac{\partial |A|}{\partial A_{nn}} \end{bmatrix}$$

$$= \begin{bmatrix} \mathrm{adj}_{11} & \cdots & 2\,\mathrm{adj}_{1n} \\ \vdots & & \\ 2\,\mathrm{adj}_{n1} & \cdots & \mathrm{adj}_{nn} \end{bmatrix}$$

Substracting and adding adj_{ii} at entry A at position (i, i), $\forall\, i = 1$ to n

$$= \begin{bmatrix} 2adj_{11} - adj_{11} & \cdots & 2adj_{1n} \\ \vdots & & \\ 2adj_{n1} & \cdots & 2adj_{nn} - adj_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} 2adj_{11} & \cdots & 2adj_{1n} \\ \vdots & & \\ 2adj_{n1} & \cdots & 2adj_{nn} \end{bmatrix} - \begin{bmatrix} adj_{11} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & adj_{nn} \end{bmatrix}$$

$$\frac{\partial |A|}{\partial A} = 2\begin{bmatrix} adj_{11} & \cdots & adj_{1n} \\ \vdots & & \\ adj_{n1} & \cdots & adj_{nn} \end{bmatrix} - \begin{bmatrix} adj_{11} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & adj_{nn} \end{bmatrix}$$

$$\therefore \frac{\partial \log|A|}{\partial A} = \frac{1}{|A|}\left[2\, adj\, A - diag\,(adj\, A)\right]$$

$$= \frac{2\, adj A}{|A|} - diag \frac{(adj\, A)}{|A|}$$

$$= 2A^{-1} - diag\, A^{-1} \qquad \left[\because A^{-1} = \frac{adj A}{|A|}\right]$$

$$\therefore \frac{\partial \log|A|}{\partial A} = 2A^{-1} - diag\, A^{-1}$$

Hence Proved

Question (2) (ii) To prove

$\frac{\partial}{\partial A} tr(AB) = 2B - diag(B)$ where A, B are square symmetric matrices

Let A, B be matrices of size $n \times n$

$$AB = \begin{bmatrix} \sum_{k=1}^{n} A_{1k} \cdot B_{k1} & \cdots & \sum_{k=1}^{n} A_{1k} \cdot B_{kn} \\ \vdots & & \\ \sum_{k=1}^{n} A_{nk} \cdot B_{k1} & \cdots & \sum_{k=1}^{n} A_{nk} \cdot B_{kn} \end{bmatrix}$$

$$\Rightarrow tr(AB) = \sum_{k=1}^{n} A_{1k} \cdot B_{k1} + \sum_{k=1}^{n} A_{2k} \cdot B_{k2} + \cdots + \sum_{k=1}^{n} A_{nk} \cdot B_{kn}$$

Claim

$$\frac{\partial\, tr(AB)}{\partial A_{ij}} = \begin{cases} B_{ii} & \text{if } i=j \\ 2B_{ij} & \text{if } i \neq j \end{cases}$$

Proof:

case (1) when $i=j$

other than A_{ii}, all other entries are independent of A_{ii}.

$$\therefore \frac{\partial\, tr(AB)}{\partial A_{ii}} = \frac{\partial}{\partial A_{ii}} \sum_{k=1}^{n} A_{1k} \cdot B_{k1} + \cdots + \sum_{k=1}^{n} A_{ik} \cdot B_{ki} + \cdots + \sum_{k=1}^{n} A_{nk} \cdot B_{kn}$$

Taking out A_{ii} term

$$\Rightarrow \frac{\partial}{\partial A_{ii}} \left(\sum_{k=1}^{n} A_{1k} \cdot B_{k1} + \ldots + A_{ii} \cdot B_{ii} + \sum_{\substack{k=1 \\ k \neq i}}^{n} A_{ik} \cdot B_{ki} + \ldots + \sum_{k=1}^{n} A_{nk} B_{kn} \right)$$

Differentiating w.r.t A_{ii}

$$= 0 + \ldots + 1 . B_{ii} + 0 + \ldots + 0$$

$$= B_{ii}$$

$$\frac{\partial tr(AB)}{\partial A_{ii}} = B_{ii} \longrightarrow (1)$$

case (2) when $i \neq j$

other than A_{ij} and A_{ji}, all other entries are independent of A_{ij}.

$$\therefore \frac{\partial tr(AB)}{\partial A_{ij}} = \frac{\partial}{\partial A_{ij}} \left(\sum_{k=1}^{n} A_{1k} B_{k1} + \ldots + \sum_{k=1}^{n} A_{ik} B_{ki} + \ldots + \sum_{k=1}^{n} A_{jk} B_{kj} + \ldots + \sum_{k=1}^{n} A_{nk} \cdot B_{kn} \right)$$

Taking out terms with A_{ij} and A_{ji}

$$= \frac{\partial}{\partial A_{ij}} \left(\sum_{k=1}^{n} A_{1k} B_{k1} + \ldots + A_{ij} B_{ji} + \sum_{\substack{k=1 \\ k \neq j}}^{n} A_{ik} B_{ki} + \ldots + A_{ji} B_{ij} + \sum_{\substack{k=1 \\ k \neq i}}^{n} A_{jk} B_{kj} + \ldots + \sum_{k=1}^{n} A_{nk} \cdot B_{kn} \right)$$

Differentiating w.r.t A_{ij}

$= 0 + \dots + 1.B_{ji} + 0 + \dots 1.B_{ij} + 0 + \dots + 0$

$= B_{ji} + B_{ij}$

$= B_{ij} + B_{ij} = 2B_{ij}$ [$\because$ B is symmetric]

$\therefore \frac{\partial tr(AB)}{\partial A_{ij}} = 2B_{ij} \rightarrow (2)$

Hence proved the claim

$$\frac{\partial tr(AB)}{\partial A} = \begin{bmatrix} \frac{\partial tr(AB)}{\partial A_{11}} & \dots & \frac{\partial tr(AB)}{\partial A_{1n}} \\ \vdots & & \vdots \\ \frac{\partial tr(AB)}{\partial A_{n1}} & \dots & \frac{\partial tr(AB)}{\partial A_{nn}} \end{bmatrix}$$

$$= \begin{bmatrix} B_{11} & \dots & 2B_{1n} \\ \vdots & & \\ 2B_{n1} & \dots & B_{nn} \end{bmatrix}$$

For all $i = 1$ to n, adding and substracting B_{ii}

$$= \begin{bmatrix} 2B_{11} - B_{11} & \dots & 2B_{1n} \\ \vdots & & \vdots \\ 2B_{n1} & \dots & 2B_{nn} - B_{nn} \end{bmatrix}$$

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

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

$$= 2B - diag(B)$$

$$\therefore \frac{\partial tr(AB)}{\partial A} = 2B - diag(B)$$

Hence proved

Question (3) (a)

To show: $S_T^y = I$, where I is $d \times d$ Identity matrix

Given

$$y_n = \Lambda^{-1/2} W^T (x-\mu)$$

$$S_T^y = \frac{1}{N} \sum_{n=1}^{N} y_n y_n^T$$

Substituting value of y_n

$$\Rightarrow S_T^y = \frac{1}{N} \sum_{n=1}^{N} (\Lambda^{-1/2} W^T (x-\mu)) (\Lambda^{-1/2} W^T (x-\mu))^T$$

As $\Lambda^{-1/2}$ is a diagonal matrix,

$$\Lambda^{-1/2} = (\Lambda^{-1/2})^T$$

$$\therefore S_T^y = \frac{1}{N} \sum_{n=1}^{N} (\Lambda^{-1/2} W^T (x-\mu))((x-\mu)^T W \Lambda^{-1/2})$$

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

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

w.k.t $S_x = \frac{1}{N} \sum_{n=1}^{N} (x-\mu)(x-\mu)^T$

$$\therefore S_T^y = \Lambda^{-1/2} W^T S_x W \Lambda^{-1/2}$$

w.k.t $S_x W = \Lambda W$ where, W is the eigen vectors of S_x with eigen values Λ

$\Rightarrow \Lambda^{-1/2} W^T \Lambda W \Lambda^{-1/2}$

$= \Lambda^{-1/2} \Lambda W^T W \Lambda^{-1/2}$

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

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

$= \Lambda^0 I \Lambda^0$

$= I$

$\therefore S_T^y = I$

Hence proved

Question (3) (b)

To show : The first LDA projection vector w is given by the eigenvector of S_W^y with minimum magnitude of eigen value.

Proof :

The first LDA projection vector w is given by the eigen vector of $S_W^{y-1} \cdot S_B^y$ with maximum eigen value,

where,

S_W^y = Within - class scatter matrix of data points $y_1 \cdots y_N$

S_B^y = Between - class scatter matrix

w.k.t $S_B^y = S_T^y - S_W^y$

As $S_T^y = I$

$\Rightarrow S_B^y = I - S_W^y$

$\therefore$ Find eigen vector of $(S_W^y)^{-1} \cdot (I - S_W^y)$ with emaximum eigen value.

Let $S_W^y \cdot w' = \lambda w'$

i.e let w' be the eigen vector of S_W^y with eigen value λ.

Then $(S_W^y)^{-1} \cdot (I - S_W^y)$

$$= (S_w^y)^{-1} \cdot I - (S_w^y)^{-1}(S_w^y)$$

$$= (S_w^y)^{-1} - I$$

As λ is the eigen value of S_w^y for the eigen vector w'

$$\therefore ((S_w^y)^{-1} - I) w' = \left(\frac{1}{\lambda} - 1\right) w'$$

To maximize $\frac{1}{\lambda} - 1$

$$\Rightarrow \max_{w'} \left(\frac{1}{\lambda} - 1\right)$$

$$\Rightarrow \max_{w'} \left(\frac{1-\lambda}{\lambda}\right)$$

$$\Rightarrow \min_{w'} \left(\frac{\lambda}{1-\lambda}\right)$$

As S_w^y is a positive semi-definite matrix

$\lambda \geq 0$.

As $\lambda \to 0$, $1-\lambda \to 1$

$$\therefore \min_{w'} \frac{\lambda}{1} = \min_{w'} \lambda$$

Hence shown, that the first LDA projection vector w^* is given by the eigenvector w' of S_w^y with minimum magnitude of eigen value λ.