Name: Asadullah Hill Galib

CSE 847

HW 2

Linear Algebra

(1)

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Here, $|A - \lambda I| = 0$

$$\Rightarrow \begin{vmatrix} 2-\lambda & 1 & 0 \\ 1 & 2-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = 0$$

$$\Rightarrow (2-\lambda)(2-3\lambda+\lambda^2-0) - 1(1-\lambda-0) + 0(0-0) = 0$$

$$\Rightarrow (4-8\lambda+5\lambda^2-\lambda^3) - 1(1-\lambda) = 0$$

$$\Rightarrow -\lambda^3+5\lambda^2-7\lambda+3 = 0$$

$$\Rightarrow -(\lambda-1)(\lambda^2-4\lambda+3) = 0$$

$$\Rightarrow -(\lambda-1)^2(\lambda-3) = 0$$

$$\Rightarrow \lambda - 1 = 0 \quad \text{or} \quad \lambda - 3 = 0$$

$$\Rightarrow \lambda = 1 \quad \text{or} \quad \lambda = 3$$

$$\therefore \lambda = 1, 3$$

For $\lambda = 1$,

$$A - \lambda I = A - I = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

P.T.O.

Using Gaussian Elimination processe:

$$A-\lambda I = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad [R_2 \leftarrow R_2 - R_1]$$

So, $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$\Rightarrow x_1 + x_2 = 0 \quad \Rightarrow x_1 = -x_2$

Eigen vector, $v = \begin{bmatrix} -x_2 \\ x_2 \\ x_3 \end{bmatrix}$

Let, $x_2 = 0$, $x_3 = 1$, then $v_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

Let, $x_2 = 1$, $x_3 = 0$, then $v_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$

for $\lambda = 3$

$$A - \lambda I = A - 3.I = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

Using Gaussian elimination process:

$$A-\lambda I = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \quad [R_1 \leftarrow R_1 / (-1)]$$

$$= \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{bmatrix} \quad [R_2 \leftarrow R_2 - R_1]$$

$$= \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & -2 \\ 0 & 0 & 0 \end{bmatrix} \quad [\text{swapping } R_2 \Leftrightarrow R_3]$$

$$= \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad [R_2 \leftarrow R_2 / (-2)]$$

So, $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$\Rightarrow x_1 - x_2 = 0$ and $x_3 = 0$

$\Rightarrow x_1 = x_2$

Eigenvector, $v = \begin{bmatrix} x_2 \\ x_2 \\ 0 \end{bmatrix}$

Let, $x_2 = 1$, $v_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ (Ans).

$\lambda = 1, 3$

$v_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ $v_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ $v_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ (Ans.)

(2) $v_1 = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}$ $\quad v_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$, $\quad v_3 = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}$

v_1, v_2 & v_3 form an orthogonal set, iff, their standard euclidian in[ner] product is equal t[o] to zero

$$\Rightarrow v_1^T . v_2 = v_2^T . v_3 = v_3^T . v_1 = 0$$

$$v_1^T . v_2 = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}^T . \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = 2\times 0 + 0(-1) + (-1)\times 0 = 0$$

$$v_2^T . v_3 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}^T . \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix} = 0\times 2 + (-1)0 + 0(4) = 0$$

$$v_3^T v_1 = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}^T . \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = 2\times 2 + 0\times 0 + 4(-1) = 0$$

So, they form orthogonal set

But, they don't form an orthonormal set as

$$\|v_1\|_2 = \sqrt{2^2 + 0^2 + (-1)^2} = \sqrt{5} \neq 1,$$

$$\|v_2\|_2 = \sqrt{0^2 + (-1)^2 + 0^2} = 1, \quad \text{[it is okay]}$$

$$\|v_3\| = \sqrt{2^2 + 0^2 + 4^2} = \sqrt{20} \neq 1$$

$\cancel{\|v\|}$ We can convert/turn them into a set of vectors that will form an orthonormal set of vectors under the standard euklidian inner product for $\mathbb{R}^3$ by normalizing,

$$u_1 = \frac{v_1}{\|v_1\|_2} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{5} \\ 0 \\ -1/\sqrt{5} \end{bmatrix}$$

$$u_2 = \frac{v_2}{\|v_2\|_2} = \frac{v_2}{1} = v_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

$$u_3 = \frac{v_3}{\|v_3\|_2} = \frac{1}{\sqrt{20}}\begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{20} \\ 0 \\ 4/\sqrt{20} \end{bmatrix} \quad [\sqrt{20} = 4\sqrt{5}$$

Now, u_1, u_2, u_3 form an orthonormal set of vectors as $\|u_1\|_2 = \|u_2\|_2 = \|u_3\|_2 = 1$

Linear Algebra

3

$A_{n \times m}$ has linearly independent columns.

$$A = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$$

$x_1 v_1 + x_2 v_2 + \cdots x_n v_n = 0$

$\Rightarrow Ax = 0 \quad \cdots\cdots (1) \qquad [\, x = [x_1, \ldots x_n]^T \,]$

$\Rightarrow x = 0$

So, $\mathcal{N}(A) = \{\vec{0}\}$ [as A's columns are linearly independent]

P.T.O.

Let, $x \in \mathcal{N}(A^T A)$, ~~then~~,

$\Rightarrow (A^T A)x = 0$

$\Rightarrow x^T A^T A x = 0$

$\Rightarrow (Ax)^T Ax = 0$

$\Rightarrow \|Ax\|^2 = 0$

$\Rightarrow Ax = 0 \Rightarrow x = 0$

or $x \in N(A)$, ~~So x is also~~

So, ~~x~~ if $x \in \mathcal{N}(A^T A)$ then $x \in \mathcal{N}(A)$

~~$\Rightarrow x = \vec{0}$~~

$\therefore \mathcal{N}(A^T A) = \mathcal{N}(A) = \{0\}$

Therefore, the columns of $A^T A$ are linearly independent and $\underbrace{A^T A}_{\underbrace{m \times n \quad n \times m}_{m \times m}}$ is a square matrix.

$\therefore A^T A$ is invertible.

(A) A is a $n \times m$ matrix with linearly independent columns.

In the question no. (3) we have shown that, if $A_{n \times m}$ has ~~with~~ linearly independent columns, $A^T A$ is an invertible matrix.

So there is a least square solution $\bar{x}$ and its associated normal system,

$$A^T A \bar{x} = A^T b$$

As $A^T A$ is invertible,

$$(A^T A)^{-1} A^T A \bar{x} = (A^T A)^{-1} A^T b$$

$$I \bar{x} = (A^T A)^{-1} A^T b$$

$$\bar{x} = (A^T A)^{-1} A^T b$$

As inverse of a matrix is unique,

$\bar{x}$ is the unique solution to the system of equation $Ax = b$. [Showed]

(2) $$p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N) \quad \ldots (3.49)$$

$$\mathbf{m}_N = \mathbf{S}_N(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\Phi\mathbf{t}) \quad \ldots (3.50)$$

$$\mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta\Phi^T\Phi \quad \ldots (3.51)$$

From Bayes's theorem, $p(\mathbf{w}|\mathbf{t}) \propto p(\mathbf{t}|\mathbf{w})\, p(\mathbf{w})$

$$\Rightarrow \cancel{p(\mathbf{w}|\mathbf{t})} \propto \left[\prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^T\phi(x_n), \beta^{-1})\right] \mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0)$$

$$\propto \exp\left(-\frac{\beta}{2}(\mathbf{t} - \Phi\mathbf{w})^T(\mathbf{t} - \Phi\mathbf{w})\right) \exp\left(-\frac{1}{2}(\mathbf{w}-\mathbf{m}_0)^T\mathbf{S}_0^{-1}(\mathbf{w}-\mathbf{m}_0)\right)$$

$$\propto \exp\left(+\frac{\beta}{\cancel{2}}\left(\mathbf{t}^T\cancel{\mathbf{t}} - \cancel{2}\mathbf{w}^T\Phi^T\right)(\mathbf{t} - \Phi\mathbf{w})\right\} + \cancel{\tfrac{1}{2}}(\mathbf{w}^T - \mathbf{m}_0^T)\mathbf{S}_0^{-1}(\mathbf{w} - \mathbf{m}_0)\Big)$$

[omitting $-\frac{1}{2}$ & expanding transpose]

$$\propto \exp\left(\beta\mathbf{t}^T\mathbf{t} - 2\beta\mathbf{w}^T\Phi^T\mathbf{t} + \beta\mathbf{w}^T\Phi^T\Phi\mathbf{w} + \mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{w} \right.$$
$$\left. \cancel{-2\mathbf{m}_0} - 2\mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{m}_0 + \mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0\right)$$

$$\propto \exp\Big(\beta\mathbf{w}^T\Phi^T\Phi\mathbf{w} + \mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{w} - \cancel{2}\big(2\mathbf{w}^T(\beta\Phi^T\mathbf{t} + \mathbf{S}_0^{-1}\mathbf{m}_0\big) +$$
$$\underbrace{\beta\mathbf{t}^T\mathbf{t} + \mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0}_{\text{constant}}\Big)$$

$$\propto \exp\Big(\mathbf{w}^T(\beta\Phi^T\Phi + \mathbf{S}_0^{-1})\mathbf{w} - 2\mathbf{w}^T\cdot\mathbf{S}_N^{-1}\mathbf{S}_N(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\Phi^T\mathbf{t})$$
$$+ \text{constant}\Big)$$

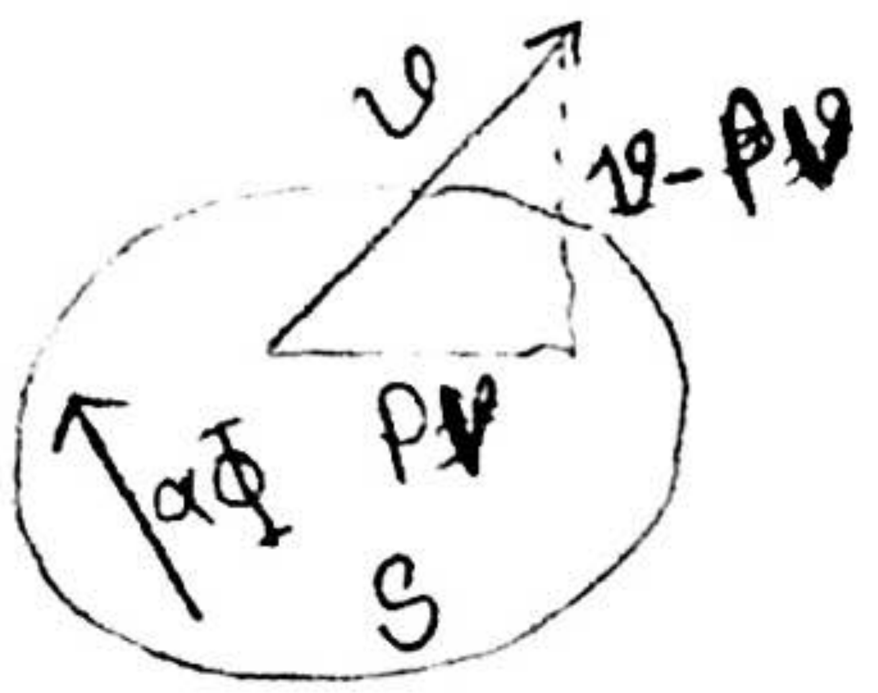
$$\propto \exp\left(\mathbf{w}^T\mathbf{S}_N^{-1}\mathbf{w} - 2\mathbf{w}^T\mathbf{S}_N^{-1}\mathbf{m}_N + \text{constant}\right)$$

$$\propto \exp\left((\mathbf{w} - \mathbf{m}_N)^T\mathbf{S}_N^{-1}(\mathbf{w} - \mathbf{m}_N)\right) \quad [\mathbf{m}_N^T\mathbf{S}_N^{-1}\mathbf{m}_N = \text{constant}]$$

$$\propto \mathcal{N}(\omega \mid m_N, S_N) = P(\omega \mid \mathcal{D})$$ [Showed]

Linear Regression

(1)



Here, $P = \Phi(\Phi^T\Phi)^{-1}\Phi^T$

$$Pv = \Phi(\Phi^T\Phi)^{-1}\Phi^T v$$

$$v - Pv = v(I - P) = v(I - \Phi(\Phi^T\Phi)^{-1}\Phi^T)$$

Let an arbitrary vector $\alpha\Phi$ lies in the subspace S.

Now,

$$(v - Pv).\alpha\Phi$$

$$= \alpha.v(I - \Phi(\Phi^T\Phi)^{-1}\Phi^T).\Phi$$

$$= \alpha.v(\cancel{\Phi} I.\Phi - \Phi\underbrace{(\Phi^T\Phi)^{-1}.\Phi^T.\Phi}_{I})$$

$$= \alpha.v(\Phi - \Phi.I)$$

$$= \alpha.v(\Phi - \Phi)$$

$$= 0$$

As their dot product is 0, so matrix P projects v onto the space S. [Showed]

P.T.O.

Now, $w_{ML} = (\Phi^T \Phi)^{-1} \Phi^T t$

w_{ML} is a vector. Because,

let Φ is $n \times m$ matrix & t is a $n \times 1$ vector.

$$\text{then, } w_{ML} = \underbrace{\underbrace{\underbrace{(\underbrace{\Phi^T \Phi}_{m\times n \; n\times m})^{-1}}_{m\times m} \underbrace{\Phi^T}_{m\times n}}_{m\times n} \underbrace{t}_{n\times 1}}_{m\times 1}$$

From

$y = \Phi w_{ML}^{*} = \Phi (\Phi^T \Phi)^{-1} \Phi^T t$

So This is an orthogonal projection of t

onto the manifold S, as,

Φw_{ML} evaluates to a ~~vectora~~ vector & we can

(Matrix. Vector $\rightarrow$ vector)

think of it as a linear combination of

the columns of Φ. So, according to the

first part, w_{ML} corresponds to an

orthogonal projection of t on the S.

[Showed

③ Using (3·3), (3·8) & (3·49), we can rewrite (3·5) as

$$p(t|x,\mathbf{t},\alpha,\beta) = \int p(t|x,w,\beta)\, p(w|\mathbf{t},\alpha,\beta)\, dw$$

$$= \int \underbrace{\mathcal{N}(t|\phi(x)^T w, \beta^{-1})}_{\text{matching with (2·114)}} \underbrace{\mathcal{N}(w|m_N, S_N)}_{\text{matching with (2·113)}} dw$$

matching with (2·114)

$\Downarrow$

$t = y$

$\phi(x)^T = A$

$w = x$

$\beta^{-1} = L^{-1}$

matching with (2·113)

$\Downarrow$

$w = x$

$m_N = \mu$

$S_N = \Lambda^{-1}$

Using (2·115), $[\, p(y) = \mathcal{N}(y|A\mu + b, L^{-1} + A\Lambda^{-1}A^T)$

$$p(t|x,\mathbf{t},\alpha,\beta) = \mathcal{N}\left(t \mid \phi(x)^T m_N,\ \beta^{-1} + \phi(x)^T S_N \phi(x)\right)$$

$$= \mathcal{N}\left(t \mid m_N^T \phi(x)^T,\ \frac{1}{\beta} + \phi(x)^T S_N \phi(x)\right) \quad \text{---}(3·58)$$

and, $\sigma_N^2(x) = \frac{1}{\beta} + \phi(x)^T S_N \phi(x)$

(4) $$(M + vv^T)^{-1} = M^{-1} - \frac{(M^{-1}v)(v^T M^{-1})}{1 + v^T M^{-1} v} \quad \ldots (3.110)$$

From (3.59)

$$\sigma^2_{N+1} = \frac{1}{\beta} + \phi(x)^T S_{N+1} \phi(x) \quad \ldots (3.59)$$

Using P (r.f. Question 3.8):

$$S_{N+1}^{-1} = S_N^{-1} + \beta \phi_{N+1} \phi_{N+1}^T \quad \ldots (P)$$

Using (3.110) & $\cancel{(3.59)}$ (P)

$$S_{N+1} = \left[S_N^{-1} + \cancel{\beta\phi(x_{N+1}} \, \beta \phi_{N+1} \phi_{N+1}^T \right]^{-1}$$

$$= S_N - \frac{\cancel{S}(\sqrt{\beta}\, S_N \phi_{N+1}).(\sqrt{\beta}\, \cancel{S_N} \phi_{N+1}^T S_N)}{1 + \beta \phi_{N+1}^T S_N \phi_{N+1}}$$

$$= S_N - \frac{\beta S_N \phi_{N+1} \phi_{N+1}^T S_N}{1 + \beta \phi_{N+1}^T S_N \phi_{N+1}}$$

Incorporating $\cancel{(P)}$ it into (3.59),

$$\sigma^2_{N+1}(x) = \frac{1}{\beta} + \phi(x)^T S_{N+1} \phi(x)$$

$$= \cancel{\sigma^2_N(x)} + \frac{1}{\beta} + \phi(x)^T \left(S_N - \frac{\beta S_N \phi_{N+1} \phi_{N+1}^T S_N}{1 + \beta \phi_{N+1}^T S_N \phi_{N+1}} \right) \phi(x)$$

$$\cancel{= \sigma^2_N(x) - \beta \phi(x)^T S_N \phi(x_{N+1}) \phi(x_{N+1})^T}$$

$$= \sigma^2_N(x) - \frac{\beta \phi(x)^T S_N \phi(x_{N+1}) \phi(x_{N+1})^T S_N \phi(x)}{1 + \beta \phi(x_{N+1})^T S_N \phi(x_{N+1})}$$

$$\sigma^2_{N+1}(x) = \sigma^2_N(x) - \frac{\phi(x)^T S_N \phi(x_{N+1}) \phi(x_{N+1})^T S_N \phi(x)}{\frac{1}{\beta} + \phi(x_{N+1})^T S_N \phi(x_{N+1})}$$

$$= \sigma^2_N(x) - \underbrace{\frac{\cancel{\phi(x)^T S_N \phi(x_{N+1})}\left(\phi(x)^T S_N \phi(x_{N+1})\right)^2}{\frac{1}{\beta} + \phi(x_{N+1})^T S_N \phi(x_{N+1})}}$$

or equal to (≥ 0)

→ this part is larger ~~than~~ 0, because the numerator ≥ 0 [square] & denominator is greater than 0 . and S_N is positive definite matrix.

So, $\sigma^2_{N+1}(x) \leq \sigma^2_N(x)$