Amirali Omidfar UID: 204869241

1

HW 1 Amirali Omidfar

1 a)

i) $A = \begin{bmatrix} .6 & .8 \\ .8 & -.6 \end{bmatrix}$ $AA^T = \begin{bmatrix} .6 & .8 \\ .8 & -.6 \end{bmatrix}\begin{bmatrix} .6 & .8 \\ .8 & -.6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ✓

$AX = \lambda X$

$AX - \lambda X = 0$

$(A - I\lambda)X = 0$

$\begin{bmatrix} .6-\lambda & .8 \\ .8 & -0.6-\lambda \end{bmatrix} X = 0$

$\det|A - \lambda I| = (.6-\lambda)(-.6-\lambda) - .8^2$

$\Rightarrow \lambda^2 - 1 = 0 \Rightarrow \lambda = \boxed{\begin{matrix} 1 \\ -1 \end{matrix}}$

if $\lambda_1 = 1 \Rightarrow \left[\begin{array}{cc|c} -.4 & .8 & 0 \\ .8 & -1.6 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cc|c} -.4 & .8 & 0 \\ 0 & 0 & 0 \end{array}\right]$

$\boxed{\text{Eigenvector } V_1 = \begin{bmatrix} x_1 \\ x_1/2 \end{bmatrix}}$

if $\lambda = -1 \Rightarrow \left[\begin{array}{cc|c} 1.6 & .8 & 0 \\ .8 & .4 & 0 \end{array}\right] \Rightarrow \left[\begin{array}{cc|c} 1.6 & .8 & 0 \\ 0 & 0 & 0 \end{array}\right]$

Eigenvector $V_2 = \begin{bmatrix} x_1 \\ -2x_1 \end{bmatrix}$

$\Rightarrow |\det A| = 1$ & eigenvectors are orthogonal to each other

$|\lambda| = 1$

ii) $Av_i = \lambda_i v_i$

$\Rightarrow \|Av_i\|_2^2 = \|\lambda_i v_i\|_2^2$

$\Rightarrow v_i^T \underbrace{A^T A}_{I} v_i = \lambda_i^2 \|v_i\|_2^2$

$\Rightarrow v_i^T v_i = \underbrace{\lambda_i^2}_{1} \|v_i\|_2^2$

$\Rightarrow \|v_i\|_2^2 = \|v_i\|_2^2$ $\checkmark$

iii) Given v_1 & v_2 being the eigenvectors for λ_1 & λ_2

$\Rightarrow v_1 . v_2 = v_2 . v_1$

$\Rightarrow A(v_1 . v_2) = A(v_2 . v_1)$

$\Rightarrow (Av_1) . v_2 = (Av_2) . v_1$

$\Rightarrow (\lambda_1 v_1) . v_2 = (\lambda_2 v_2) . v_1$

$\Rightarrow \lambda_1 (v_1 . v_2) - \lambda_2 (v_2 . v_1) = 0$

$\Rightarrow (\lambda_1 - \lambda_2)(v_1 . v_2) = 0$

since $\lambda_1 \neq \lambda_2 \Rightarrow \lambda_1 - \lambda_2 \neq 0$

So $v_1 . v_2 = 0 \Rightarrow \boxed{v_1 \perp v_2 \checkmark}$

iv) Any x is subject to rotation or reflection under Ax when $|\det A| = 1$ in this case its reflection as $\det A = -1$

The length is preserved.

b)

i) $A = U \Sigma V^T$

$$AA^T = U \Sigma \underbrace{V^T V}_{I} \Sigma^T U^T =$$

$$AA^T = U \underbrace{\Sigma \Sigma^T}_{\text{diagonal matrix}} U^T$$

$$A^T A = V \Sigma^T \underbrace{U^T U}_{I} \Sigma V^T$$

$$A^T A = V \underbrace{\Sigma^T \Sigma}_{\text{diagonal matrix}} V^T$$

According to Eigendecomposition $\Rightarrow$ singular vectors of A are eigenvectors of AA^T & $A^T A$.

Also λ's for $A^T A$ & AA^T are σ's for A (singular values)

ii) As shown ~~Above~~ $\therefore$ $AA^T = \underbrace{U}_{V} \underbrace{\Sigma \Sigma^T}_{\Lambda} U^T \Rightarrow$ so

λ's for AA^T & AA^T are singular values of A squared!!

c)

i) False $\quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ has one only

ii) False $\quad$ if eigenvalues are different then it doesn't hold

$$V_3 = V_1 + V_2 \Rightarrow A[V_3] = AV_1 + AV_2 = \lambda_1 V_1 + \lambda_2 V_2 \neq \lambda(V_1 + V_2)$$

unless $\lambda_1 = \lambda_2$

iii) True. If λ is an eigenvalue for $A \Rightarrow Ax = \lambda x$

$$\Rightarrow \begin{cases} x^T A x \geq 0 \\ x^T A x = \lambda x^T x \end{cases} \Rightarrow \begin{cases} \lambda x^T x \geq 0 \\ x^T x \geq 0 \text{ is always greater than zero} \end{cases}$$

$\Rightarrow \lambda \geq 0$ So the λ's are non negative.

iv) False, according to the theorem: $A \in R^{m \times n} \Rightarrow n = \text{Rank}A + \text{nullity of } A$

$\Rightarrow$ if v is in the nullity of $A \Rightarrow Av = 0$ so v is eigenvector for $\emptyset$ eigenvalue. Since there exists an eigenspace of degree at least one for each eigenvalue, the number of non-zero eigenvalues must be less than or equal to the rank.

v) True, Assume: $A \in \mathbb{R}^{n \times n}$ & $\begin{matrix} AV = \lambda V \\ AV' = \lambda V' \end{matrix}$

$V, V' \in \mathbb{R}^{n \times 1}$

$\Rightarrow C = A(V + V') \Rightarrow C_{i1} = \sum_{k=1}^{n} a_{ik}(v_{k1} + v'_{k1}) = \sum a_{ik} v_{k1} + \sum a_{ik} v'_{k1}$

$\Rightarrow C = AV + AV' \Rightarrow \lambda V + \lambda V' = \lambda(V + V') \Rightarrow A(V + V') = \lambda(V + V')$ //

2. $P(H|H50) = 0.5$, $P(H|H60) = 0.6$

i) $P(H50|\overset{\text{tail}}{T}) = \dfrac{P(T|H50)P(H50)}{P(T|H50)P(H50) + P(T|H60)P(H60)}$

$= \dfrac{(0.5)(0.5)}{(0.5)(0.5) + (0.4)(0.5)} = \boxed{5/9}$

ii)

$S = THHH$

$P(H50|S) = \dfrac{P(S|H50)P(H50)}{P(S|H50)P(H50) + P(S|H60)P(H60)}$

$= \dfrac{(0.5)^4(0.5)}{(0.5)^4(0.5) + (0.4)(0.6)^3(0.5)} \approx 0.4197$

iii) S = 9 heads in ten attempts

$P(S|H50) = \binom{10}{9}(0.5)^9(0.5) = .0097$

$P(S|H55) = \binom{10}{9}(0.55)^9(0.45) = 0.0207$

$P(S|H60) = \binom{10}{9}(0.6)^9(0.4) = 0.0403$

$P(Hi|S) = \dfrac{P(S|Hi)P(Hi)}{\sum_{Hi=50,55,60} P(S|Hi)P(Hi)}$

$\Rightarrow P(H50|S) = \dfrac{(0.0097)(1/3)}{0.0707(1/3)} = \boxed{0.1372}$

$P(H55|S) = \dfrac{(0.0207)(1/3)}{0.0707(1/3)} = \boxed{0.2928}$

$P(H60|S) = \dfrac{(0.0403)(1/3)}{0.0707(1/3)} = \boxed{0.5700}$

b) $P(+|Pr) = 0.99$

$P(+|NPr) = 0.1$

$P(NPr) = 0.99$

$$P(Pr|+) = \frac{P(+|Pr)\,P(Pr)}{P(+|Pr)\,P(Pr) + P(+|NPr)\,P(NPr)}$$

$$= \frac{(0.99)(0.01)}{(0.99)(0.01) + (0.1)(0.99)}$$

$$= \frac{1}{11} = 0.\overline{09}$$

It means the test is not reliable when only in 9% of the time the women are pregnant given the test being positive.

c) $E(X) = \sum_x x\,P(X)$

$\Rightarrow E(AX+b) = \sum_x (AX+b)\,P(x)$

$$= A\sum x\,P(x) + b\underbrace{\sum P(x)}_{1}$$

$$= A\sum x\,P(x) + b$$

$$= \boxed{AE(X)+b}$$

Using scaling & additivity properties

$\Rightarrow$ linearity of expection with respect to one variable

7

2.d) $\text{Cov}(x) = E((x-E(x))(x-E(x))^T)$

$$\text{Cov}(Ax+b) = E\left(\left(Ax+b-E(Ax+b)\right)\left(Ax+b-E(Ax+b)\right)^T\right)$$

$$= E\left(\left(Ax+\cancel{b}-AE(x)-\cancel{b}\right)\left(Ax+b-AE(x)-b\right)^T\right)$$

$$= E\left(\left(Ax-AE(x)\right)\left(Ax-AE(x)\right)^T\right)$$

$$= E\left(\left[A(x-E(x))\right]\left[A(x-E(x))\right]^T\right)$$

$$= E\left(A\underbrace{(x-E(x))(x-E(x))^T}_{\text{Cov}(x)}A^T\right)$$

$$\boxed{= A\,\text{Cov}(x)\,A^T}$$

3. a) $x \in R^n,\ y \in R^m,\ A \in R^{n\times m} \Rightarrow \nabla_x x^T A y?$

$f = \nabla_x x^T A y \Rightarrow \in R^{n\times 1} \mid f_i = \frac{\partial}{\partial x_i} x^T A y$

$$f = x^T A y = [x_1 \cdots x_n]\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ & \ddots & \\ a_{n1} & & a_{nm}\end{bmatrix}\begin{bmatrix} y_1 \\ \vdots \\ y_m\end{bmatrix}$$

$$f = \sum_{j=1}^{m}\left(\sum_{i=1}^{n} x_i a_{ij}\right) y_j$$

$$\frac{\partial f}{\partial x_k} = \frac{\partial \sum_{j=1}^{m}(x_k a_{kj})y_j}{\partial x_k} = \frac{\partial}{\partial x_k} x_k \sum_{j=1}^{m}(a_{kj}y_j) = \sum_{j=1}^{m} a_{kj}y_j$$

$$\Rightarrow \nabla_x f = \begin{bmatrix} \sum_{j=1}^{m} a_{1j}y_j \\ \vdots \\ \sum_{j=1}^{m} a_{nj}y_j \end{bmatrix} = \boxed{Ay}$$

$1 \times n$

b) $\nabla_y x^T A y$?

$$f = x^T A y = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} x_i a_{ij} \right) y_j$$

$$\nabla_y f = \begin{bmatrix} \sum_{i=1}^{n} x_i a_{i1} \\ \vdots \\ \vdots \\ \sum_{i=1}^{n} x_i a_{in} \end{bmatrix} = \boxed{A^T x}$$

c) $\nabla_A x^T A y$?

$$f = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} x_i a_{ij} \right) y_j$$

$$\nabla_A f = \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \cdots & \frac{\partial f}{\partial a_{1m}} \\ & & \\ \frac{\partial f}{\partial a_{n1}} & & \frac{\partial f}{\partial a_{nxm}} \end{bmatrix} \Rightarrow \frac{\partial f}{\partial a_{kl}} = x_k y_l$$

$$= \begin{bmatrix} x_1 y_1 & \cdots & x_1 y_m \\ \vdots & & \vdots \\ x_n y_1 & & x_n y_m \end{bmatrix} = \boxed{x y^T}$$

d) $f = x^T A x + b^T x$

$$\nabla_x f = \nabla_x x^T A x + \nabla_x b^T x$$

$$\nabla_x b^T x = \begin{bmatrix} \frac{\partial \sum b_i x_i}{\partial x_1} \\ \vdots \\ \frac{\partial \sum b_i x_i}{\partial x_n} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \boxed{b}$$

$$x^T A x = \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{n} x_i a_{ij} \right)$$

$\Rightarrow$ linearity of expection with respect to one variable

$$\frac{\partial x^TAx}{\partial x_k} = \frac{\sum x_j(\sum x_i a_{ij})}{\partial x_k} = \sum x_i(a_{ik}) + \sum x_i(a_{kj})$$

$$= \underline{A^Tx + Ax}$$

$$\Rightarrow \nabla_x(x^TAx + b^Tx) = \boxed{A^Tx + Ax + b}$$

e) $f = tr(AB) \Rightarrow \nabla_A f$?

$A \in R^{m\times n}$, $B \in R^{n\times m} \Rightarrow C = AB$, $C \in R^{m\times m}$

$$c_{kl} = \sum_{i=1}^{n} a_{ki} b_{il}$$

$$tr(AB) = tr(C) = \sum_{j=1}^{m} c_{jj} = \sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ji} b_{ij}\right)$$

$$\nabla_A f = \begin{bmatrix} \frac{\partial(AB)}{\partial a_{11}} & \cdots & \frac{\partial(AB)}{\partial a_{1n}} \\ \vdots & & \\ \frac{\partial(AB)}{\partial a_{m1}} & & \frac{\partial(AB)}{\partial a_{mn}} \end{bmatrix}$$

$$\Rightarrow \frac{\partial f}{\partial a_{ij}} = b_{ji} \qquad \Rightarrow \boxed{\nabla_A f = B^T}$$

4

$$\min_W \frac{1}{2}\sum_{i=1}^{n} \|y^i - Wx^i\|^2$$

$$\min_W \frac{1}{2}\sum_{i=1}^{n} (y^i - Wx^i)^T(y^i - Wx^i)$$

$$\min_W \frac{1}{2}(Y - XW^T)^T(Y - XW^T)$$

$$\min_W \frac{1}{2}(Y^T - WX^T)(Y - XW^T)$$

$$\min_W \underbrace{\frac{1}{2}(Y^TY - Y^TXW^T - WX^TY + WX^TXW^T)}_{\to L}$$

$$\nabla_W L = \frac{1}{2}(0 - 2Y^TX + 2WX^TX) = 0 \qquad WX^TX = X^TY$$

$$\boxed{W = Y^TX(X^TX)^{-1}}$$