EE 239 AS HW1

Problem 1

a) Since $AA^T = I$, the rows & cols must be mutually orthonormal → A is an orthogonal matrix.

i. let $A = \begin{bmatrix} 4/5 & 3/5 \\ 3/5 & -4/5 \end{bmatrix}$. $|A - \lambda I| = \begin{vmatrix} 4/5 - \lambda & 3/5 \\ 3/5 & -4/5 - \lambda \end{vmatrix} = 0$

$$(4/5 - \lambda)(-4/5 - \lambda) - 9/25 = 0$$

$$-\frac{16}{25} \cancel{- \frac{4}{5}\lambda + 4/5\lambda} + \lambda^2 - 9/25 = 0 \rightarrow -1 + \lambda^2 = 0,\ \lambda^2 = 1,$$

$$\boxed{\begin{matrix}\lambda_1 = 1 \\ \lambda_2 = -1\end{matrix}}\ \boxed{\text{eigenvalues}}$$

Eigenvectors:

$Av = v \rightarrow \begin{bmatrix} 4/5 & 3/5 \\ 3/5 & -4/5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$

$4/5 x + 3/5 y = x$

$\frac{3}{5}x - 4/5 y = y \rightarrow \frac{3}{5}x = 9/5 y \rightarrow y = x/3$ (x, y are elements of the eigenvector)

$\rightarrow 4/5 x + \left(\frac{3}{5}\right)(x/3) = x \rightarrow x = x \rightarrow$ unit vector s.t. $y = x/3$

$\rightarrow x^2 + \frac{x^2}{9} = 1 \rightarrow \frac{10x^2}{9} = 1,\ x = \sqrt{9/10} \rightarrow x = \frac{3}{\sqrt{10}},\ y = \frac{1}{\sqrt{10}}$

$\boxed{\text{An eigenvector corresponding to } \lambda_1 = 1 \text{ is } \begin{bmatrix} \frac{3\sqrt{10}}{10} \\ \frac{\sqrt{10}}{10} \end{bmatrix}}$

the eigenvalues are complements of each other (or complex conjugates if they were complex), the eigenvectors are normal to each other. They also are ~~a basis~~ an orthonormal basis for $\mathbb{R}$

$Av = -v$

$\rightarrow \begin{bmatrix} 4/5 & 3/5 \\ 3/5 & -4/5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x \\ -y \end{bmatrix}$

$4/5 x + 3/5 y = -x$

$3/5 x - 4/5 y = -y \rightarrow 3/5 x = -y/5,$ $3x = -y,$ ~~y=3x~~ $x = -y/3$

$\rightarrow \cancel{3/5}(-y/\cancel{3}) - \frac{4}{5}y = -y \rightarrow -y = -y \rightarrow$ unit vec s.t. $x = -y/3$

$y^2 + y^2/9 = 1 \rightarrow y = 3/\sqrt{10},\ x = -1/\sqrt{10}$

$\boxed{\text{eigenvector corresponding to } \lambda_2 = -1 \text{ is } \begin{bmatrix} \frac{-\sqrt{10}}{10} \\ \frac{3\sqrt{10}}{10} \end{bmatrix}}$

1.

a) $\text{rank}(\lambda)Ax$

~~i. Since A is an orthogonal matrix,~~

ii. We have $Av = \lambda v$. The L_2 norms on both sides must therefore be equal:

$\|Av\| = \|\lambda v\|$. Since λ is a (potentially complex) scalar, we can take it out:

→ complex modulus/norm of eigenvalue

$$\|Av\| = \|\lambda\| \|v\|$$

since $\|x\| = \sqrt{x^T x}$, we have

$$\sqrt{(Av)^T Av} = \|\lambda\| \|v\|$$

$$\sqrt{v^T A^T A v} = \|\lambda\| \|v\|$$

I

$$\sqrt{v^T I v} = \|\lambda\| \|v\|$$

$$\Rightarrow \sqrt{v^T v} = \|\lambda\| \|v\|$$

$\|v\| = \|\lambda\| \|v\|$ so $\|\lambda\| = 1$ to keep this true.

iii. We have $x^T y = x^T I y = x^T A^T A y$

$$= (Ax)^T (Ay) = (\lambda_1 x)^T (\lambda_2 y) = \lambda_1 \lambda_2 x^T y$$

so $x^T y = \lambda_1 \lambda_2 x^T y$. Assume $x^T y \neq 0$, this means that $\lambda_1 = \lambda_2$. But if $\lambda_1 \in \mathbb{R}$ and $\lambda_2 \in \mathbb{R}$, then since $\|\lambda\| = 1$ and $\lambda_1 \neq \lambda_2$ (distinct), $\lambda_1 \lambda_2 \neq 1$, so we have a contradiction and $x^T y = 0$. Now, let $\lambda_1 \in \mathbb{R}$ and $\lambda_2 \in \mathbb{C}$ or $\lambda_1 \in \mathbb{C}$ and $\lambda_2 \in \mathbb{R}$, the product of $\lambda_1 \lambda_2 \neq 1$ in this case, since one λ is complex and the other is not, so $x^T y = 0$ here also. Now, let $\lambda_1 \in \mathbb{C}$ and $\lambda_2 \in \mathbb{C}$. If λ_1 and λ_2 are not conjugates, then $\lambda_1 \lambda_2$ will be complex, so $\lambda_1 \lambda_2 \neq 1$ and $x^T y = 1$. If λ_1 and λ_2 are complex conjugates, then

iv. In general, the inner products are preserved: $(Ax)^T (Ay)$
$= x^T A^T A y = x^T I y = x^T y$, meaning that the transformation A corresponds to a rotation or reflection, meaning that the vector x may be rotated by some degree θ or reflected about some axis.

Problem 1

b)

i. Left-singular ~~values of~~ vectors of A = eigenvectors of AA^T ((citation: p.43 of DL book)

Right-singular vectors of A = eigenvectors of A^TA

ii. (Nonzero) singular values of A: $\sqrt{\text{eigenvalues}(A^TA)}$ ((citation: p.43 of DL book)

$= \sqrt{\text{eigenvalues}(AA^T)}$

(SVD: $A = UDV^T$ where cols of U = left-singular vectors of A, cols of V = right singular vectors, and D = diag matrix whose diagonals are singular values of A.

c.

i. False, there's at most n distinct values

ii. False: If $Av_1 = \lambda_1 v_1$ & $Av_2 = \lambda_2 v_2$, then $A(v_1+v_2) = Av_1 + Av_2 = \lambda_1 v_1 + \lambda_2 v_2$ generally $\neq \lambda_n (v_1+v_2)$.

iii. True. We have $x^TAx \geq 0 \to x^T\lambda x \geq 0$

$\to \lambda x^T x \geq 0$, since $x \neq \vec{0}$,

$\lambda \geq 0$ ✓

iv. False. By rank-nullity,

we have $\text{rank}(A) + \text{nullity}(A) = n$.

If $\text{nullity}(A) = x > 0$, then there are $n-x$ ~~eigen~~ nonzero eigenvalues, and $\text{rank}(A) = n - \text{nullity}(A) = n-x$.

v. True,

If $Av_1 = \lambda v_1$ & $Av_2 = \lambda v_2$

$A(v_1+v_2) = Av_1 + Av_2 = \lambda v_1 + \lambda v_2 = \lambda(v_1+v_2)$ ✓

Problem 2

a)

i. $p(H50 \mid tails) = \dfrac{p(tails \mid H50)\, p(H50)}{p(tails)} = \dfrac{\left(\frac{1}{2}\right)^2}{p(T, H50) + p(T, H60)}$

$= \dfrac{\left(\frac{1}{2}\right)^2}{p(T \mid H50)\, p(H50) + p(T \mid H60)\, p(H60)} = \dfrac{\left(\frac{1}{2}\right)^2}{\left(\frac{1}{2}\right)^2 + \left(\frac{2}{5}\right)\left(\frac{1}{2}\right)} = \dfrac{\frac{1}{4}}{\frac{1}{4} + \frac{2}{5}} = \dfrac{\left(\frac{1}{4}\right)}{\left(\frac{9}{20}\right)} = \boxed{5/9}$

ii. $p(H50 \mid THHH)$

independence of flips

$= \dfrac{p(THHH \mid H50)\, p(H50)}{p(THHH)} = \dfrac{p(T \mid H50)\left(p(H \mid H50)\right)^3 p(H50)}{p(THHH \mid H50)\, p(H50) + p(THHH \mid H60)\, p(H60)}$

$= \dfrac{p(T \mid H50)\, p(H \mid H50)\, p(H50)}{p(T \mid H50)\, p(H \mid H50)^3\, p(H50) + p(T \mid H60)\, p(H \mid H60)^3\, p(H60)}$

independence of flips

$= \dfrac{(0.5)(0.5)^3(0.5)}{\frac{1}{2^5} + \left(\frac{2}{5}\right)\left(\frac{3}{5}\right)^3\left(\frac{1}{2}\right)} = \dfrac{\frac{1}{2^5}}{\frac{1}{2^5} + \left(\frac{1}{5}\right)\left(\frac{27}{125}\right)}$

$= \dfrac{\left(\frac{1}{32}\right)}{\left(\frac{1}{32} + \frac{27}{625}\right)} = \boxed{0.41974}$

Problem 2

a)

iii. $p(H50 \mid H_9T_1) = \dfrac{p(H_9T_1 \mid H50)\, p(H50)}{p(H_9T_1)}$

$$= \frac{(0.5)^{10}\left(\frac{1}{3}\right)}{\sum p(H_9T_1 \mid i)\, p(i)} = \frac{(0.5)^{10}\left(\frac{1}{3}\right)}{(0.5)^{10}\left(\frac{1}{3}\right) + \left(\frac{11}{20}\right)^9\left(\frac{9}{20}\right)\left(\frac{1}{3}\right) + \left(\frac{12}{20}\right)^9\left(\frac{8}{20}\right)\left(\frac{1}{3}\right)}$$

$i: H_{50}, H_{55}, H_{60}$

$$= \frac{\left(\frac{1}{2^{10}}\right)\left(\frac{1}{3}\right)}{\left[\left(\frac{1}{2^{10}}\right)\left(\frac{1}{3}\right) + \left(\frac{11}{20}\right)^9\left(\frac{9}{20}\right)\left(\frac{1}{3}\right) + \left(\frac{3}{5}\right)^9\left(\frac{2}{5}\right)\left(\frac{1}{3}\right)\right]}$$

define as $p(H_9T_1) \leftarrow$

$\hookrightarrow = 0.00236$

$$= \frac{\left(\frac{1}{3072}\right)}{p(H_9T_1)} = \boxed{0.1379 = p(H50 \mid H_9T_1)}$$

$$p(H_{55} \mid H_9T_1) = \frac{p(H_9T_1 \mid H_{55})\, p(H_{55})}{p(H_9T_1)} = \frac{\left(\frac{11}{20}\right)^9 \left(9/20\right)\left(1/3\right)}{p(H_9T_1)}$$

$$= \boxed{0.2927 = p(H_{55} \mid H_9T_1)}$$

$$p(H_{60} \mid H_9T_1) = \frac{p(H_9T_1 \mid H_{60})\, p(H_{60})}{p(H_9T_1)} = \frac{\left(\frac{3}{5}\right)^9 \left(2/5\right)\left(1/3\right)}{p(H_9T_1)}$$

$$\boxed{= 0.5694 = p(H_{60} \mid H_9T_1)}$$

check: $0.1379 + 0.2927 + 0.5694 = 1$ ✓

Problem 2

b) define 1 = test positive, p = pregnant.

- $p(1|p) = 0.99$
  $p(1|\sim p) = 0.1$
  $p(p) = 0.01$

- find $p(p|1)$.

$$p(p|1) = \frac{p(1|p)\,p(p)}{p(1)}$$

$$p(1) = p(1,p) + p(1,\sim p)$$
$$= p(1|p)p(p) + p(1|\sim p)p(\sim p)$$

$$= \frac{p(1|p)p(p)}{p(1|p)p(p) + p(1|\sim p)p(\sim p)}$$

$$= \frac{(0.99)(0.01)}{(0.99)(0.01) + (0.1)(0.99)} \quad \boxed{= 0.0909}$$

This surprisingly low probability does make sense if we consider the high false positive rate, 10%. This would mean that 10% of 99% of the population would get a false positive, a pretty large amount. If our false positive rate were lower, such as 0.1%, then our probability would go up to about 91%.

c) $E[Ax+b] = E[Ax] + E[b] = E[Ax] + b$
$\boxed{= AE[x] + b}$ (due to linearity of expectations)

d) $\mathrm{cov}(Ax+b) = E((Ax+b - E(Ax+b))(Ax+b - E(Ax+b))^T)$
$= E((Ax+b - (AE(x)+b))(Ax+b - (AE(x)+b))^T)$
$= E((Ax - AE(x))(Ax - AE(x))^T) = E(A(x-E(x))(A(x-E(x)))^T)$
$= E(A(x-E(x))(x-E(x))^T A^T) \rightarrow$ linearity of expect $\rightarrow AE((x-E(x))(x-E(x))^T)A^T$
$\boxed{= A\,\mathrm{cov}(x)A^T}$

$2^5 \quad (5)(5) \; \bar{} \chi)$

$$\frac{\frac{1}{25}+\left(\frac{1}{5}\right)\left(\frac{27}{125}\right)}{} = \frac{(3^2)}{\frac{1}{32}+\frac{27}{625}} = 0.4197$$

Problem 3

3) a) $\nabla_x x^T A y$

$\cancel{= Ay + x^T A(0)} = \boxed{Ay}$

$\nabla_x x^T A y = \nabla_x (Ay)^T x = \boxed{Ay}$

b) $\nabla_y x^T A y = (x^T A)^T = \boxed{A^T x}$

c) $\nabla_A x^T A y = \cancel{x^T y + x y^T = 2 x y^T}$

$\nabla_A \; x^T A y = xy^T$

$$x^T A y = \sum_{i=1} \sum_{j=1} x_i y_j a_{ij} \rightarrow \nabla_{a_{ij}} \sum_{i=1} \sum_{j=1} x_i y_j a_{ij} \rightarrow x_i y_j$$

$\boxed{= xy^T}$ (matches A's dim)

Problem 3

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

$\nabla_x x^T A x$

$\rightarrow \sum_i \sum_j x_i A_{ij} x_j$

$\frac{d}{dx_1} \cdot \sum_i \sum_j x_i A_{ij} x_j$

when $i,j=1 \rightarrow x_1 x_1 A_{11}$

$\rightarrow 2 A_{11} x_1$ is the derivative

when $i=1, j\neq 1$, $x_1 A_{1j} x_j$

$d/dx = A_{1j} x_j$

when $i \neq 1, j=1$, $\frac{d}{dx_1}(x_i A_{i1} x_1) = x_i A_{ij}$

$\rightarrow \sum_{j\neq 1} A_{ij} x_j + \sum_{i \neq 1} x_i A_{ij} + 2A_{11}x_1$

$\rightarrow \sum_j A_{ij} x_j + \sum_i x_i A_{ij}$ is the general term,

so we have $Ax + A^T x$

$\nabla_x b^T x \rightarrow \sum_i b_i x_i \rightarrow \frac{d}{dx_i} = b_i$, so $\nabla_x b^T x = b$

$$\boxed{\nabla_x f = Ax + A^T x + b}$$

~~e) let C = AB. Then~~ $\cancel{C_{ij} = \sum_k A_{nk} B_{kn}} \rightarrow \cancel{\nabla_{A_{nk}} \sum_k A_{nk} B_{kn} = B_{kn}}$

~~so each element is~~ $\cancel{B_{kn}} \rightarrow \cancel{\boxed{B^T}}$

~~(A is 20~~ ~~(dim of B is also~~ ~~width dim of A)~~

e) let $C = AB$. Then

$C_{ij} = \sum_k A_{ik} B_{kj}$. Since we're taking

$tr(C)$, we have the diagonal elements

of C: $C_{nn} = \sum_k A_{nk} B_{kn}$.

Now, $\nabla_{A_{nk}} \sum_k A_{nk} B_{kn} = B_{kn}$, so each

element will be B_{kn}, so

$$\nabla_A tr(AB) = \boxed{B^T}$$

4)

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

(Frobenius norm, from book)

First, replace $\|A\|^2$ w/ $tr(AA^T)$:

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

$$= \frac{1}{2}\sum_{i=1}^{n} tr[\,y^i y^{iT} \underbrace{- y_i x^{iT} W^T - W x^i y^{iT}}_{\text{combine}} + Wx^i(Wx^i)^T\,]$$

$$= \frac{1}{2}\sum_{i=1}^{n} tr[\,y^i y^{iT} - 2Wx^i y^{iT} + Wx^i(Wx^i)^T\,]$$

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

$$\frac{d}{dW}\left(\quad \ldots \quad \right)$$

$$= \frac{d}{dW}\left(-\sum_{i=1}^{n} tr(Wx^i y^{iT})\right) + \frac{d}{dW}\left(\frac{1}{2}\sum_{i=1}^{n} tr(Wx^i(Wx^i)^T)\right)$$

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

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

$$\rightarrow \sum_{i=1}^{n}(x^i y^{iT})^T = \sum_{i=1}^{n} Wx^i x^{iT}$$

$$\sum_{i=1}^{n}(x^i y^{iT})^T = W\sum_{i=1}^{n} x^i x^{iT}$$

$$\rightarrow \sum_{i=1}^{n} y^i x^{iT} = W\sum_{i=1}^{n} x^i x^{iT}$$

$$\rightarrow \boxed{\hat{W} = \left(\sum_{i=1}^{n} x^{iT} x^i\right)^{-1}\sum_{i=1}^{n} y^i x^{iT}}$$

~~or contracting~~
or using ~~matr~~ X = design matrix where row i is the i^{th} example, and y is a vector of labels, we have

$$\boxed{\hat{W} = (X^T X)^{-1} X^T y}$$