HW 1

① (a)(i) $AA^T = I$, where A is a 2×2 square matrix

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, so $A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$

Then $AA^T = \begin{bmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{bmatrix}$, which must be I

→ So we get the following criteria:

$ac+bd = 0 \qquad a^2+b^2 = c^2+d^2 = 1$

Then a suitable choice for A would be:

$A = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ Now finding eigenvalues & eigenvectors:

$\det(A-\lambda I) = 0$:

$\det\left(\begin{bmatrix} -\frac{1}{\sqrt{2}}-\lambda & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}-\lambda \end{bmatrix}\right) = \left(-\frac{1}{\sqrt{2}}-\lambda\right)\left(\frac{1}{\sqrt{2}}-\lambda\right) - \left(\frac{1}{\sqrt{2}}\right)^2$

$= \lambda^2 - 1 = 0 \rightarrow \boxed{\lambda_1 = 1,\ \lambda_2 = -1}$

Finding eigenvectors $v_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ for λ_1:

$(A-\lambda_1 I)\vec{v}_1 = 0$

$\begin{bmatrix} -\frac{1}{\sqrt{2}}-\lambda_1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}-\lambda_1 \end{bmatrix}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} \frac{-1-\sqrt{2}}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1-\sqrt{2}}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$-\left(\frac{1+\sqrt{2}}{\sqrt{2}}\right)x_1 + \frac{1}{\sqrt{2}}y_1 = 0 \qquad \frac{1}{\sqrt{2}}x_1 + \frac{(1-\sqrt{2})}{\sqrt{2}}y_1 = 0$

$y_1 = (1+\sqrt{2})x_1 \qquad x_1 = (\sqrt{2}-1)y_1$

So $\vec{v}_1$ is of the form $k_1\begin{bmatrix} 1 \\ 1+\sqrt{2} \end{bmatrix}$, where k_1 is a const.

① ⓐ (i) (continued)

$(A - \lambda_2 I)\vec{v}_2 = 0$

$$\begin{bmatrix} -\frac{1}{\sqrt{2}} - \lambda_2 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} - \lambda_2 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{\sqrt{2}} + 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} + 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \begin{matrix} (1) \\ (2) \end{matrix}$$

$$\frac{(\sqrt{2} - 1)}{\sqrt{2}} x_2 + \frac{1}{\sqrt{2}} y_2 = 0 \qquad (1)$$

$$\rightarrow y_2 + (\sqrt{2} - 1) x_2 = 0$$

$$y_2 = (1 - \sqrt{2}) x_2$$

$$\frac{1}{\sqrt{2}} x_2 + \frac{1 + \sqrt{2}}{\sqrt{2}} y_2 = 0$$

$$\rightarrow x_2 + (1 + \sqrt{2}) y_2 = 0$$

$$x_2 = -(1 + \sqrt{2}) y_2 \qquad (2)$$

$$y_2 = -\frac{1}{1 + \sqrt{2}} x_2 \cdot \frac{1 - \sqrt{2}}{1 - \sqrt{2}}$$

$$= -\frac{(1 - \sqrt{2})}{-1} x_2$$

$$= (1 - \sqrt{2}) x_2$$

So we get $\vec{v}_2 = k_2 \begin{bmatrix} 1 \\ 1 - \sqrt{2} \end{bmatrix}$

The magnitudes of the eigenvalues are 1 and the eigenvectors are orthogonal to each other

① ⓐ(ii) $A v_i = \lambda_i v_i$

$$\|A v_i\|^2 = \|\lambda_i v_i\|^2$$

$$v_i^T \|A\|^2 v_i = \|\lambda_i\|^2 \|v_i\|^2$$

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

$$v_i^T v_i = \|\lambda_i\|^2 \|v_i\|^2$$

$$\rightarrow \boxed{\|\lambda_i\|^2 = 1}$$

(iii) By commutativity of dot product, $v_1 \cdot v_2 = v_2 \cdot v_1$

Multiply both sides by A:

$$A v_1 \cdot v_2 = A v_2 \cdot v_1 \quad (eq\ 1)$$

Since $A\vec{v}_1 = \lambda_1 \vec{v}_1$ and $A\vec{v}_2 = \lambda_2 \vec{v}_2$, eq 1 becomes

$$\lambda_1 v_1 \cdot v_2 = \lambda_2 v_2 \cdot v_1$$

$$\lambda_1 v_1 \cdot v_2 - \lambda_2 v_2 \cdot v_1 = 0$$

$$\lambda_1 v_1 \cdot v_2 - \lambda_2 v_1 \cdot v_2 = 0$$

$$v_1 \cdot v_2 (\lambda_1 - \lambda_2) = 0$$

because $\lambda_1 \neq \lambda_2$, $v_1 \cdot v_2$ must be zero;

therefore, $\boxed{v_1 \perp v_2}$

(iv) If $\vec{v}$ is an eigenvector of A, then applying A to v will have the same result as multiplication by the corresponding eigenvalue.

If $\vec{v}$ is some linear combo of eigenvectors of A, then because they're orthogonal, the respective components of each eigenvector will be scaled by the corresponding eigenvalues.

① ⓑ (i) Singular value decomposition of A:

$$A = U \Sigma V^T \qquad A^T = V \Sigma^T U^T$$

So $AA^T = U \Sigma V^T V \Sigma^T U^T$

$= U \Sigma \Sigma^T U^T$, which is a square matrix

Which means the eigenvalues and eigenvectors of AA^T would be the singular values and singular vectors of A.

$A^T A = V \Sigma^T U^T U \Sigma V^T = V \Sigma^T \Sigma V^T$, so the same would apply to $A^T A$.

(ii) The singular values are the eigenvalues of $A^T A$ and AA^T, arranged in descending order.

① ⓒ (i) Two eigenvalues could be the same, False

(ii) Using $A\vec{v} = \lambda\vec{v}$:

$A(v_1+v_2) = Av_1 + Av_2$

$Av_1 + Av_2 = \lambda_1 v_1 + \lambda_2 v_2$ because of

but since $\lambda_1 \neq \lambda_2$, we can't factor, so False

(iii) True

(iv) ~~[illegible]~~ True

(v) $Av_1 + Av_2 = \lambda v_1 + \lambda v_2$ (assume $A \in \mathbb{R}^{2\times 2}$)

$A(v_1+v_2) = \lambda(v_1+v_2)$

by definition, (v_1+v_2) corresponds to an eigenvalue of A, True

② ⓐ (iii) Now we have $P(H50) = P(H55) = P(H60) = \frac{1}{3}$

10 flips $\rightarrow$ Geom $(k=9)$

Let this sequence be known as B.

$P(B|H50) = (0.5)(0.5)^9$

$P(B|H55) = (0.45)(0.55)^9$

$P(B|H60) = (0.4)(0.6)^9$

$$P(B) = P(B|H50)P(H50) + P(B|H55)P(H55) + P(B|H60) \cdot P(H60)$$
$$= (0.5)^{10} \cdot \frac{1}{3} + (0.45)(0.55)^9 \cdot \frac{1}{3} + (0.4)(0.6)^9 \cdot \frac{1}{3}$$
$$\approx 0.00236$$

$$P(H50|B) = \frac{P(B|H50) \cdot P(H50)}{P(B)} \approx \boxed{0.137931}$$

$$P(H55|B) = \frac{P(B|H55) \cdot P(H55)}{P(B)} \approx \boxed{0.2927}$$

$$P(H60|B) = \frac{P(B|H60) \cdot P(H60)}{P(B)} \approx \boxed{0.5694}$$

These add up to 1 so solution makes sense

(2) (a) (i) Select coin from jar $\rightarrow P(H50) = \frac{1}{2} = P(H60)$

Find $P(H50 \mid T)$.

We know $P(T \mid H50) = 0.5$. By Bayes' Rule,

$$P(H50 \mid T) = \frac{P(T \mid H50) \cdot P(H50)}{P(T)} = \frac{0.5 \cdot 0.5}{P(T)}$$

By law of total probability,

$$P(T) = P(T \mid H50) \cdot P(H50) + P(T \mid H60) \cdot P(H60)$$

So,

$$P(H50 \mid T) = \frac{P(T \mid H50) \cdot P(H50)}{P(T \mid H50) \cdot P(H50) + P(T \mid H60) \cdot P(H60)}$$

$$= \frac{(0.5)(0.5)}{(0.5)(0.5) + (0.4)(0.5)} = \frac{0.25}{0.45} \approx \boxed{0.556}$$

(ii) Replaced coin, so next coin picked still has $P(H50) = P(H60) = \frac{1}{2}$

Define sequence $\{T, H, H, H\}$ as the event A.

Assuming we don't care about order,

$$P[A \mid H50] = (0.5)^1 \cdot (0.5)^3 = 0.5^4 = \frac{1}{16} \approx 0.0625$$

$$P[A \mid H60] = (0.4)^1 \cdot (0.6)^3 = \frac{2 \cdot 3^3}{5^4} = \frac{54}{625} \approx 0.0864$$

Find $P(H50 \mid A)$:

$$P(H50 \mid A) = \frac{P(A \mid H50) \cdot P(H50)}{P(A)} = \frac{\left(\frac{1}{16} \cdot \frac{1}{2}\right)}{P(A)}$$

$$P(A) = P(A \mid H50) \cdot P(H50) + P(A \mid H60) \cdot P(H60)$$

$$= \frac{1}{16} \cdot \frac{1}{2} + \frac{54}{625} \cdot \frac{1}{2} \approx 0.07445$$

$\rightarrow P(H50 \mid A) = \boxed{0.419745}$

(This result makes sense because the sequence has more heads. Therefore you'd be inclined to say the coin was weighted towards H).

2 (b) $P(1 \mid \text{pregnant}) = 0.99$ $\qquad$ $P(\text{not preg}) = 0.99$

$P(1 \mid \text{not preg}) = 0.10$ $\qquad$ $\to P(\text{preg}) = 0.01$

Find $P(\text{preg} \mid 1)$: By Bayes Rule,

$$P(\text{preg} \mid 1) = \frac{P(1 \mid \text{preg}) \cdot P(\text{preg})}{P(1)}$$

By Total Probability,

$$P(1) = P(1 \mid \text{preg})P(\text{preg}) + P(1 \mid \text{not preg})P(\text{not preg})$$

$$= (0.99)\cdot(0.01) + (0.10)\cdot(0.99) = 0.1089$$

$$\to P(\text{preg} \mid 1) = \frac{0.99 \cdot 0.01}{0.1089} \simeq \boxed{0.09091}$$

This makes sense because the overall likelihood of being pregnant is quite low, so absent any other information (is the woman trying to get pregnant or is she, for example, not sexually active?) the test has low reliability in practice when used on "random" women. If you select from a population of women who are trying to get pregnant, P(preg | positive) is much higher for the same test, with fewer false positives.

(2) (c)

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$ x_i's ~ iid, A & b deterministic

$E(A\vec{x} + b) = E(A\vec{x}) + E(b)$, by linearity of expectation

$= E(A\vec{x}) + b$

$= \sum_{j=1}^{k} A \cdot x_j p_j + b$

$= A \sum_{j=1}^{k} x_j p_j + b$

$= A\,E[\vec{x}] + b$

(2) (d) $\text{cov}(Ax+b) = E\left[\left((Ax+b) - E(Ax+b)\right)\left((Ax+b) - E(Ax+b)\right)^T\right]$

$= E\left[(Ax+b - AE(x) - b)(Ax+b - AE(x) - b)^T\right]$

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

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

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

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

(3) ⓐ $\nabla_x \underset{(1\times n)}{x^T} \underset{(n\times m)}{A} \underset{(m\times 1)}{y}$ First ~~take~~ multiply Ay:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m} a_{1j} y_j \\ \sum_{j=1}^{m} a_{2j} y_j \\ \vdots \\ \sum_{j=1}^{m} a_{nj} y_j \end{bmatrix}$$

So $x^T(Ay) = [x_1 \ \ x_2 \ \ x_3 \dots x_n] \begin{bmatrix} \sum_{j=1}^{m} a_{1j} y_j \\ \sum_{j=1}^{m} a_{2j} y_j \\ \vdots \\ \sum_{j=1}^{m} a_{nj} y_j \end{bmatrix}$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i a_{ij} y_j$$

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

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

$$= \begin{bmatrix} \sum_{j=1}^{m} a_{1j} y_j \\ \sum_{j=1}^{m} a_{2j} y_j \\ \vdots \\ \sum_{j=1}^{m} a_{nj} y_j \end{bmatrix} = \boxed{Ay}$$

③ ⓑ Find $\nabla_y x^T A y$

from (a), we have

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

$$\nabla_y f = \begin{bmatrix} \sum_{i=1}^{n} x_i a_{i1} \\ \sum_{i=1}^{n} x_i a_{i2} \\ \vdots \\ \sum_{i=1}^{n} x_i a_{im} \end{bmatrix} = \boxed{A^T x}$$, which is $(m \times 1)$

$$= \begin{bmatrix} (x_1 a_{11} + x_2 a_{21} + \cdots + x_n a_{n1}) \\ (x_1 a_{12} + x_2 a_{22} + \cdots + x_n a_{n2}) \\ \vdots \\ (x_1 a_{1m} + x_2 a_{2m} + \cdots + x_n a_{nm}) \end{bmatrix}$$

$$= \underset{(m \times n)}{\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix}} \underset{(n \times 1)}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}$$

(3) (c) Find $\nabla_A x^T A y$. From (a), we know that:

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

$$\nabla_A f = \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \cdots & \frac{\partial f}{\partial a_{1m}} \\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} & \cdots & \frac{\partial f}{\partial a_{2m}} \\ \vdots & & & \\ \frac{\partial f}{\partial a_{n1}} & \frac{\partial f}{\partial a_{n2}} & \cdots & \frac{\partial f}{\partial a_{nm}} \end{bmatrix}$$

$$= \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_m \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_m \end{bmatrix} \quad \left(\text{which is } (n \times m)\right)$$

$$= \boxed{x\, y^T} \quad (n \times m)$$

3 d Find $\nabla_x f$, where $f = x^T A x + b^T x$

$$\nabla_x (x^T A x + b^T x) = \nabla_x (x^T A x + \nabla_x (b^T x)$$

$$\underline{\nabla_x (\cancel{\ } x^T A x)}: \quad f_1(x) = x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} x_i x_j$$

- for $i=1, j=1$, the term is $a_{11} x_1^2$

$$\frac{\partial}{\partial x_1}(a_{11} x_1^2) = 2 a_{11} x_1$$

- for $i=1, j \neq 1$, the term is $a_{1j} x_1 x_j \rightarrow \frac{\partial}{\partial x_1}(a_{1j} x_1 x_j) = a_{1j} x_j$
- and $i \neq 1, j=1$, the term is $a_{i1} x_i x_1 \rightarrow \frac{\partial}{\partial x_1}(a_{i1} x_i x_1) = a_{i1} x_i$

$$\rightarrow \text{So} \quad \frac{\partial f(x)}{\partial x_1} = 2 a_{11} x_1 + \sum_{j=2}^{m} a_{1j} x_j + \sum_{i=2}^{n} a_{i1} x_i$$

$$= \sum_{j=1}^{m} a_{1j} x_j + \sum_{i=1}^{n} a_{i1} x_i = (A\bar{x})_1 + (A^T \bar{x})_1$$

$$\rightarrow \frac{\partial f_1(x)}{\partial x} = (A + A^T) x$$

$$\nabla_x (b^T x) = \begin{bmatrix} \frac{\partial \sum_{i=1}^{n} b_i x_i}{\partial x_1} \\ \frac{\partial \sum_{i=1}^{n} b_i x_i}{\partial x_2} \\ \vdots \\ \frac{\partial \sum_{i=1}^{n} b_i x_i}{\partial x_n} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = b$$

$$\rightarrow \nabla_x (x^T A x + b^T x) = \boxed{(A + A^T) x + b}$$

③ ⓔ ~~[illegible]~~ $f = tr(AB)$, find $\nabla_A f$

A is $n \times m$, so to make AB square, B must be $m \times n$

$$AB = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ a_{21} & \ddots & \vdots \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} \sum_i a_{1i} b_{i1} & \cdots & \sum_i a_{1i} b_{in} \\ \vdots & \ddots & \vdots \\ \sum_i a_{ni} b_{i1} & \cdots & \sum_i a_{ni} b_{in} \end{bmatrix} \quad (n \times n)$$

~~call this matrix C~~

$$Tr(AB) = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ji} b_{ij}$$

$$\nabla_A Tr(AB) = \begin{bmatrix} \frac{\partial}{\partial a_{11}} \sum\sum a_{ji} b_{ij} & \cdots & \frac{\partial}{\partial a_{1m}} \sum_{i,j}\sum a_{ji} b_{ij} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial a_{n1}} \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ji} b_{ij} & \cdots & \frac{\partial}{\partial a_{nm}} \sum\sum a_{ji} b_{ij} \end{bmatrix}$$

$$= \begin{bmatrix} b_{11} & b_{21} & \cdots & b_{m1} \\ \vdots & & \ddots & \vdots \\ b_{1n} & \cdots & & b_{mn} \end{bmatrix} = \boxed{B^T}$$

(4) $\min_{W} \frac{1}{2} \sum_{i=1}^{n} \| y^{(i)} - W x^{(i)} \|^2$

$= \min_{W} \frac{1}{2} \| Y - WX \|_F^2$, where $Y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{pmatrix}$ and $X = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{pmatrix}$

$= \min_{W} \frac{1}{2} \mathrm{Tr}\left[(Y - WX)^T (Y - WX) \right]$

$= \min_{W} \frac{1}{2} \mathrm{Tr}\left[Y^T Y - Y^T W X - X^T W^T Y + X^T W^T W X \right]$

$= \min_{W} \frac{1}{2} \left[\mathrm{Tr}(Y^T Y) - \mathrm{Tr}(Y^T W X) - \mathrm{Tr}(X^T W^T Y) + \mathrm{Tr}(X^T W^T W X) \right]$

$= \min_{W} \frac{1}{2} \left[\mathrm{Tr}(Y^T Y) - 2\mathrm{Tr}(W X Y^T) + \mathrm{Tr}(X^T W^T W X) \right]$

Now, take $\frac{\partial}{\partial W}$ and set $= 0$:

$\frac{1}{2} \left(-2 Y X^T + 2 W X X^T \right) = 0$

$-Y X^T + W X X^T = 0$

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