1.

a.

a. $A=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$

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

$\left|\frac{1}{\sqrt{2}}\begin{bmatrix}\lambda & 1\\ 1 & -\lambda\end{bmatrix}-\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right|=0$

$\left|\begin{bmatrix}\frac{1}{\sqrt{2}}\lambda-1 & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{-\lambda}{\sqrt{2}}-1\end{bmatrix}\right|=0$

$\left(\frac{\lambda}{\sqrt{2}}-1\right)\left(\frac{-\lambda}{\sqrt{2}}-1\right)-\frac{1}{2}$

$\frac{-\lambda^2}{2}+\frac{\lambda}{\sqrt{2}}-\frac{\lambda}{\sqrt{2}}+1-\frac{1}{2}=0$

$\frac{-\lambda^2}{2}+\frac{1}{2}=0$

$\lambda^2=1$

$\lambda_1=1$

$\lambda_2=-1$

$Av=\lambda v$

$Av-\lambda v=0$

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

$\lambda=1$ $\begin{bmatrix}\frac{1}{\sqrt{2}}-1 & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}-1\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix}=0$

$\begin{array}{cc} a & b \end{array}$

$\left[\begin{array}{cc|c}\frac{1-\sqrt{2}}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ \frac{1}{\sqrt{2}} & \frac{-1-\sqrt{2}}{\sqrt{2}} & 0\end{array}\right]$

$\left[\begin{array}{cc|c}\frac{1-\sqrt{2}}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ \frac{-(1-\sqrt{2})}{2} & -\frac{1}{\sqrt{2}} & 0\end{array}\right]$

$-(1-\sqrt{2})(1+\sqrt{2})$
$-(1-2)$
$+1$

$\left[\begin{array}{cc|c}\frac{1-\sqrt{2}}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ 0 & 0 & 0\end{array}\right]$

$b=K_1$

$\frac{1-\sqrt{2}}{\sqrt{2}}a+\frac{1}{\sqrt{2}}K_1=0$

$a=-\frac{1}{\sqrt{2}}K_1\cdot\frac{\sqrt{2}}{1-\sqrt{2}}$

$a=\frac{-K_1}{1-\sqrt{2}}$

$e_1=K_1\begin{bmatrix}1+\sqrt{2}\\ 1\end{bmatrix}$

$\lambda = -1$

$$\left[\begin{array}{cc|c} \frac{-1-\sqrt{2}}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1-\sqrt{2}}{\sqrt{2}} & 0 \end{array}\right]$$

$$\left[\begin{array}{cc|c} \frac{-1-\sqrt{2}}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{+1+\sqrt{2}}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{array}\right]$$

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

$$\left[\begin{array}{cc|c} \frac{-1-\sqrt{2}}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 \end{array}\right]$$

$b = k_2$

$$\frac{-(1+\sqrt{2})}{\sqrt{2}} a + \frac{1}{\sqrt{2}} k_2 = 0$$

$$\frac{+(1+\sqrt{2})}{\sqrt{2}} a = \frac{1}{\sqrt{2}} k_2$$

$$a = \frac{k_2}{1+\sqrt{2}} = -k_2 \frac{(1-\sqrt{2})}{1}$$

$$e_2 = k_2 \begin{bmatrix} 1-\sqrt{2} \\ 1 \end{bmatrix}$$

$\lambda_1 = 1$ $\qquad e_1 = \begin{bmatrix} 1+\sqrt{2} \\ 1 \end{bmatrix} \cdot \frac{1}{\sqrt{1+(1+\sqrt{2})^2}}$

$\lambda_2 = -1$ $\qquad e_2 = \begin{bmatrix} 1-\sqrt{2} \\ 1 \end{bmatrix} \cdot \frac{1}{\sqrt{1+(1+\sqrt{2})^2}}$

The eigenvalues are same magnitude (magnitude of 1) and the eigenvectors are orthogonal

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

ii.

$\lambda_1 = 1 \quad |\lambda_1| = 1$

$\lambda_2 = -1 \quad |\lambda_2| = 1$

$Av = \lambda v$

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

$||Av||^2 = ||\lambda v||^2$

$(Av)^T(Av) = (\lambda v)^T(\lambda v)$

$v^T A^T A v = v^T \lambda^T \lambda v$

$v^T I v = v^T ||\lambda||^2 v$

$v^T v = ||\lambda||^2 ||v||^2$

$||v||^2 = ||\lambda||^2 ||v||^2$

$||\lambda||^2 = 1$

$\boxed{||\lambda|| = 1}$

iii. $e_1^T e_2 = 0$ if orthogonal and distinct eig values

$[1+\sqrt{2} \quad 1]\begin{bmatrix} 1-\sqrt{2} \\ 1 \end{bmatrix} = (1+\sqrt{2})(1-\sqrt{2}) + 1 = -1 + 1 = 0$

$AA^T = I \qquad U = \begin{bmatrix} \uparrow & \uparrow \\ \frac{e_1}{||e_1||} & \frac{e_2}{||e_2||} \\ \downarrow & \downarrow \end{bmatrix} \quad \Sigma = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$

$(U\Sigma U^T)(U\Sigma U^T)^T = I$

$U\Sigma\Sigma^T U^T = I$

$UIU^T = I$

$UU^T = I$

$$\begin{bmatrix} \frac{e_1^T e_1}{||e_1||^2} & \frac{e_2^T e_1}{||e_2|| \, ||e_1||} \\ \frac{e_1^T e_2}{||e_1|| \, ||e_2||} & \frac{e_2^T e_1}{||e_2|| \, ||e_1||} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$\frac{e_2^T e_1}{||e_2|| \, ||e_1||} = 0$

$\boxed{e_2^T e_1 = 0}$

iv. A vector x under the transformation A is only rotated and reflected as the magnitude of x will not change due to the norm of the eigenvalues being 1.

b. $A = U\Sigma V^T$

i. The columns of U are called the left singular vectors of A (and are orthonormal eigenvectors of AA^T).

The columns of V are called the right singular vectors of A (and are orthonormal eigenvectors of A^TA).

proof: $A \in \mathbb{R}^{m\times n}$ $U \in \mathbb{R}^{m\times 1}$ $V \in \mathbb{R}^{n\times 1}$ $eig(\underbrace{A^TA}_{\Sigma^T}) = eig(\underbrace{AA^T}_{\Sigma})$ $\Sigma^T = \Sigma$

$A^TAv_i = \sigma_i^2 v_i$

assume v_i is a unit norm eigenvector of A^TA

$v_i^TA^TAv_i = \sigma_i^2 v_i^Tv_i$

$(Av_i)^T(Av_i) = \sigma_i^2 \|v_i\|^2$

$\|Av_i\|^2 = \sigma_i^2$

$\|Av_i\| = \sigma_i$

$A^TAv_i = \sigma_i^2 v_i$

$AA^TAv_i = \sigma_i^2 Av_i$

Av_i is an eigenvector of AA^T

$u_i = \frac{Av_i}{\|Av_i\|} = \frac{Av_i}{\sigma_i}$

$Av_i = \sigma_i u_i$

$AA^Tu_i = \sigma_i^2 u_i$

assume u_i is unit norm eigenvector of AA^T,

$u_i^TAA^Tu_i = \sigma_i^2 u_i^Tu$

$(A^Tu_i)(A^Tu_i) = \sigma_i^2 \|u_i\|^2$

$\|A^Tu_i\|^2 = \sigma_i^2$

$AA^Tu_i = \sigma_i^2 u_i$

$A^TAA^Tu_i = A^T\sigma_i^2 u_i$

A^Tu_i is eigenvector of A^TA

$v_i = \frac{A^Tu_i}{\|A^Tu_i\|} = \frac{A^Tu_i}{\sigma_i}$

$v_i\sigma_i = A^Tu_i$

$u_i^TA = \sigma_i v_i^T$

$u_i^TAv_i = \sigma_i v_i^Tv_i$

$u_i^TAv_i = \sigma_i\|v_i\|^2 = \sigma_i$

$u_i^TAv_i = \sigma_i \rightarrow U^TAV = \Sigma \rightarrow \boxed{A = \underbrace{U}_{m\times m}\underbrace{\Sigma}_{m\times n}\underbrace{V^T}_{n\times n}}$

$U^TU = I$

$VV^T = I$

$\Sigma_{ii} = \sigma_i$

$\Sigma_{ij} = 0$ for $i \neq j$

ii.

$A^TA = (U\Sigma V^T)^T(U\Sigma V^T)$

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

$= V\Sigma^T\Sigma V^T$

$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T$

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

$= U\Sigma\Sigma^T U$

$$\Sigma^T\Sigma = \Sigma^T\Sigma = \begin{bmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & & \\ & & \ddots & & \\ & & & \sigma_r^2 & \\ & & & & \ddots 0 \end{bmatrix}$$

hence $\boxed{\lambda_i(AA^T) = \lambda_i(A^TA) = \sigma_i^2(A)}$

c.

i. False (it has at most n-distinct eigenvalue)

ii. False

$A(\underbrace{c_1e_1 + c_2e_2}_{v}) = c_1Ae_1 + c_2Ae_2$

$= c_1\lambda_1 e_1 + c_2\lambda_2 e_2$

iii. True.

iv. False

v. True

$A(\underline{c_1e_1 + c_2e_2}) = c_1\lambda e_1 + c_2\lambda e_2$

$= \lambda(\underline{c_1e_1 + c_2e_2})$

2.

a.

$P(H50) = .5 = P(H60)$

$P(H \mid H50) = .5 \quad P(H \mid H60) = .6$

i.

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

$$= \frac{(.5)(.5)}{(.5)(.5) + (.4)(.5)} = \boxed{0.5556}$$

ii.

$P((T,H,H,H) \mid H50) \rightarrow$ conditionally ind $= (.5)^4 \quad P((T,H,H,H) \mid H60) = (.4)(.6)^3$

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

$$= \frac{(.5)^4(.5)}{(.5)^4(.5) + (.4)(.6)^3(.5)}$$

$$\boxed{= 0.4197}$$

iii.

$$P(H50 \mid \text{the flips}) = \frac{(.5^9)(.5)(\frac{1}{3})}{(.5^9)(.5)(\frac{1}{3}) + (.6)^9(.4)(\frac{1}{3}) + (.55)^9(.45)(\frac{1}{3})} = \boxed{0.1379 = P(H50 \mid \text{the flips})}$$

$$P(H55 \mid \text{the flips}) = \frac{(.55^9)(.45)(\frac{1}{3})}{\text{same den}} = \boxed{0.2927 = P(H55 \mid \text{the flips})}$$

$$\boxed{P(H60 \mid \text{the flips}) = 0.5694}$$

b.

$P(+|\text{preg}) = .99$

$P(+|\text{not preg}) = .10$

$P(\text{not preg}) = .99$

$$P(\text{preg}|+) = \frac{P(+|\text{preg})P(\text{preg})}{P(+|\text{preg})P(\text{preg}) + P(+|\text{not preg})P(\text{not preg})}$$

$$= \frac{(.99)(.01)}{(.99)(.01) + (.10)(.99)} = \boxed{0.0909}$$

This makes sense because the prior for a woman being pregnant any time is very low (.01), but once the woman gets a positive test we receive new information and we may update the prior ~~with~~ to a posterior (shes pregnant given we know she tested positive). Now its shown ~~the~~ shes much more likely to be pregnant than she was before the test.

c. $\boxed{E(Ax+b) = \int (Ax+b)\,dX = A\int x\,dX + \underbrace{\int b\,dX}_{\text{integ of constant}} = A\,E[X]+b = A\mu_x + b}$

d. $\text{cov}(Ax+b) = E[(Ax+b-A\mu_x-b)(Ax+b-A\mu_x-b)^T]$

$E[(Ax-A\mu_x)(Ax-A\mu_x)^T]$

$E[A(x-\mu_x)(A(x-\mu_x))^T]$

$A\,E[(x-\mu_x)(x-\mu_x)^T]A^T$

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

3. $A = \begin{bmatrix} - a_1 - \\ - a_2 - \\ \vdots \\ - a_n - \end{bmatrix}$

a.

$$x^T \begin{bmatrix} a_1 y \\ a_2 y \\ \vdots \\ a_n y \end{bmatrix} = \sum_{i=1}^{n} x_i a_i y$$

$$\boxed{\nabla_x x^T A y = \begin{bmatrix} a_1 y \\ a_2 y \\ \vdots \\ a_n y \end{bmatrix} = Ay}$$

b. $A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_m \\ | & | & & | \end{bmatrix}$

$$[x a_1 \; x a_2 \cdots x a_m] y = \sum_{i=1}^{m} (x a_i) y_i \qquad \boxed{\nabla_y x^T A y = A^T x}$$

c. $\sum_{i=1} \sum_{j=1} x_i y_j a_{ij}$ $\qquad \boxed{\nabla_A x^T A y = x y^T}$

d. $\boxed{\nabla_x f = Ax + A^T x + b = (A + A^T) x + b}$

use notes in class to get gradient of $x^T A x$

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

e. $f = tr(\underset{n\times m}{A}\underset{m\times n}{B})$

$$tr\left(\begin{bmatrix} - a_1 - \\ - a_2 - \\ \vdots \\ - a_n - \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_n \\ | & | & & | \end{bmatrix} \right) = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n$$

$$\sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ji}$$

$$\boxed{\nabla_A f = B^T}$$

4.

$Z = y - Wx$ vector

$f = z^T z$ scalar

Tr(scalar) = scalar

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

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

$$\frac{1}{2}\sum_{i=1}^{n} \left(y^{(i)T} y^{(i)} - 2y^{(i)T} W x^{(i)} + x^{(i)T} W^T W x^{(i)}\right)$$

$$\frac{1}{2}\sum_{i=1}^{n} \left(y^{(i)T} y^{(i)} - 2\,\mathrm{Tr}\left(y^{(i)T} W x^{(i)}\right) + \mathrm{Tr}\left(x^{(i)T} W^T W x^{(i)}\right)\right)$$

$$f = \frac{1}{2}\sum_{i=1}^{n} \left(y^{(i)T} y^{(i)} - 2\,\mathrm{Tr}\left(W x^{(i)} y^{(i)T}\right) + \mathrm{Tr}\left(W x^{(i)} x^{(i)T} W^T\right)\right)$$

$$\frac{d}{dW} f = 0$$

$$\frac{1}{2}\sum_{i=1}^{n} \left(-2\, y^{(i)} x^{(i)T} + W(x^{(i)} x^{(i)T}) + W(x^{(i)} x^{(i)T})\right) = 0$$

$$\frac{1}{2}\sum_{i=1}^{n} \left(-2y^{(i)} x^{(i)T} + 2W(x^{(i)} x^{(i)T})\right) = 0$$

$$\sum_{i=1}^{n} \left[W(x^{(i)} x^{(i)T}) - y^{(i)} x^{(i)T}\right] = 0$$

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

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

can make as matrices as well ... $\boxed{W = (Y^T X)(X^T X)^{-1}}$

$$X = \begin{bmatrix} - x_1^T - \\ - x_2^T - \\ \vdots \\ - x_n^T - \end{bmatrix} \qquad Y = \begin{bmatrix} - y_1^T - \\ - y_2^T - \\ \vdots \\ - y_n^T - \end{bmatrix}$$