EE239AS
HW1
Yutong Lu
005027184

1. Linear algebra refresher

(a) i. Let $A = \begin{bmatrix} \frac{4}{5} & \frac{3}{5} \\ -\frac{3}{5} & \frac{4}{5} \end{bmatrix}$ satisfying $AA^T = I$.

$Av = \lambda v \Rightarrow (A - \lambda I)v = 0 \Rightarrow \det(A - \lambda I) = 0$

$\Rightarrow \det \begin{bmatrix} \frac{4}{5} - \lambda & \frac{3}{5} \\ -\frac{3}{5} & \frac{4}{5} - \lambda \end{bmatrix} = 0 \Rightarrow (\frac{4}{5} - \lambda)^2 - \frac{3}{5}(-\frac{3}{5}) = 0$

$\Rightarrow \frac{16}{25} + \lambda^2 - \frac{8}{5}\lambda + \frac{9}{25} = 0 \Rightarrow \lambda^2 - \frac{8}{5}\lambda + 1 = 0 \Rightarrow \lambda_1 = 0.8 + 0.6i, \ \lambda_2 = 0.8 - 0.6i$

When $\lambda_1 = 0.8 + 0.6i$,

$\begin{bmatrix} \frac{4}{5} - \frac{4}{5} - \frac{3}{5}i & \frac{3}{5} \\ -\frac{3}{5} & \frac{4}{5} - \frac{4}{5} - \frac{3}{5}i \end{bmatrix} v_1 = 0 \Rightarrow \begin{bmatrix} -\frac{3}{5}i & \frac{3}{5} \\ -\frac{3}{5} & -\frac{3}{5}i \end{bmatrix} v_1 = 0 \Rightarrow v_1 = \begin{bmatrix} -\frac{\sqrt{2}}{2}i \\ \frac{\sqrt{2}}{2} \end{bmatrix}$

when $\lambda_2 = 0.8 - 0.6i$,

$\begin{bmatrix} \frac{4}{5} - \frac{4}{5} + \frac{3}{5}i & \frac{3}{5} \\ -\frac{3}{5} & \frac{4}{5} - \frac{4}{5} + \frac{3}{5}i \end{bmatrix} v_2 = 0 \Rightarrow \begin{bmatrix} \frac{3}{5}i & \frac{3}{5} \\ -\frac{3}{5} & \frac{3}{5}i \end{bmatrix} v_2 = 0 \Rightarrow v_2 = \begin{bmatrix} \frac{\sqrt{2}}{2}i \\ \frac{\sqrt{2}}{2} \end{bmatrix}$

Eigenvalues can involve complex numbers rather than real numbers, and in this case, the two eigenvalues are complex conjugate to each other. and both of them have a norm of 1.
As for eigenvectors, they are also complex conjugate to each other, and $v_1^T v_2 = 0$, they are orthogonal.

ii. $AA^T = I$,

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

$\Rightarrow v^T (A^T A) v = |\lambda|^2 v v^T$

$\Rightarrow v^T v = |\lambda|^2 v v^T \Rightarrow \|v\|^2 = |\lambda|^2 \|v\|^2 \Rightarrow$ A has eigenvalues with norm 1.

iii. Assume u and v are eigenvectors corresponding to eigenvalues λ_u and λ_v of the matrix A.

$\begin{cases} Au = \lambda_u \cdot u \Rightarrow u^T A = \lambda_u \cdot u^T & ① \\ Av = \lambda_v \cdot v & ② \end{cases}$

By multiplying ① with ②, $u^T \cancel{A A}^{I} v = \lambda_u \cdot u^T \lambda_v \cdot v \Rightarrow u^T v = \lambda_u \cdot \lambda_v \cdot u^T v$

$\Rightarrow u^T v (1 - \lambda_u \cdot \lambda_v) = 0$, then $u^T v = 0$ and $\lambda_u \cdot \lambda_v = 1$ due to arbitrary

Therefore, distinct eigenvectors are orthogonal.

iv. Before multiplying by A, we have unit eigenvectors on an unit circle. By plotting A, it can be observed that A distorts the unit circle, it scales space in direction $v^{(i)}$ by λ_i.

(b) i. The left singular vectors of A are the eigenvectors of AA^T. The right singular vectors of A are the eigenvectors of $A^T A$.

ii. The non-zero singular values of A are the square roots of the eigenvalues of A^TA and it is also true for A^TA.

(c) i. FALSE. There are linear operators with no eigenvalues, and actually, it should be that there are at most n distinct eigenvalues.

ii. FALSE. For example, $A=\begin{bmatrix}1&0\\0&2\end{bmatrix}$ and $\begin{bmatrix}1\\0\end{bmatrix}$ is the eigenvector of eigenvalue 1, and $\begin{bmatrix}0\\1\end{bmatrix}$ is the eigenvector of eigenvalue 2. However, $\begin{bmatrix}1\\1\end{bmatrix}$ is not an eigenvector of A.

iii. TRUE. $x^TAx = x^TQ\Lambda Q^Tx = (Q^Tx)^T\Lambda(Q^Tx) = \sum_{i=1}^{n}\lambda_i(q_i^Tx)^2 \geq \lambda_n\sum_{i=1}^{n}(q_i^Tx)^2 = \lambda_n\|x\|^2$

Then, $\lambda_n\|x\|^2 \geq 0$, and this ensures its eigenvalues are non-negative.

iv. TRUE. Suppose $L: \mathbb{R}^n \to \mathbb{R}^m$, then $\text{rank}(L) (= \dim R(L)) + \text{nullity}(L) = m$, then the no. of non-zero eigenvalues must be less or equal to the rank of a matrix.

v. FALSE Since these two eigenvectors are corresponding to one eigenvalue, then they can be linearly independent so that their sum may not be an eigenvector.

2. (a) i. $p(\text{head}|\text{"H50"}) = 0.5$, $p(\text{head}|\text{"H60"}) = 0.6$, find $p(\text{"H50"}|\text{tail})$?

$$p(\text{"H50"}|\text{tail}) = \frac{p(\text{tail}|\text{"H50"})\,p(\text{"H50"})}{p(\text{tail})} = \frac{p(\text{tail}|\text{"H50"})\,p(\text{"H50"})}{\sum_H p(\text{tail}|\text{"H"})}$$

$$= \frac{p(\text{tail}|\text{"H50"})\,p(\text{"H50"})}{p(\text{tail}|\text{"H50"})\,p(\text{"H50"}) + p(\text{tail}|\text{"H60"})\,p(\text{"H60"})}$$

$$= \frac{0.5\times0.5}{0.5\times0.5+(1-0.6)\times0.5} = \frac{5}{9}$$

ii. If the coin is of type H50, the probability of THHH is $(0.5)^4$

If the coin is of type H60, the probability of THHH is $0.4\times(0.6)^3$

The probability of it will be H50 is:

$$\frac{(0.5)^4}{(0.5)^4+0.4\times(0.6)^3} = 0.42$$

iii If the coin is of type H50, the probability of 9 heads out of ten flips is $C_{10}^9(0.5)^9C_{10}^1(0.5)$.

If the coin is of type H55, the probability of 9 heads out of ten flips is $C_{10}^9(0.55)^9C_{10}^1(1-0.55)$

If the coin is of type H60, the probability of 9 heads out of ten flips is $C_{10}^9(0.6)^9C_{10}^1(1-0.6)$

Then, the probability of coin being H50 is $\frac{C_{10}^9(0.5)^9C_{10}^1(0.5)}{C_{10}^9(0.5)^9C_{10}^1(0.5)+C_{10}^9(0.55)^9C_{10}^1(0.45)+C_{10}^9(0.6)^9C_{10}^1(0.4)}$ bn x.

$$= \frac{(0.5)^{10}}{(0.5)^{10}+(0.55)^9(0.45)+(0.6)^9(0.4)} = 0.138$$

The probability of the coin being H55 is $\frac{(0.55)^9(0.45)}{(0.5)^{10}+(0.55)^9(0.45)+(0.6)^9(0.4)} = 0.293$

The probability of the coin being H60 is $\frac{(0.6)^9(0.4)}{(0.5)^{10}+(0.55)^9(0.45)+(0.6)^9(0.4)} = 0.569$

(b) Find $p(\text{pregnant}|+ve)$

$$p(\text{pregnant}|+ve) = \frac{p(+ve|\text{pregnant})\,p(\text{pregnant})}{p(+ve|\text{pregnant})\,p(\text{pregnant}) + p(+ve|\text{not})\,p(\text{not})}$$

$$= \frac{99\% \times 1\%}{99\% \times 1\% + 10\% \times 99\%} = 9.09\%$$

Since 99% of the woman population is not pregnant at any time point, even the test indicates positive, it may just be a false positive, and this can also be inferred by high false positive rate (=10%) given in the question.

(c) $E[Ax+b] = \sum_x (Ax+b)p(x) = \sum_x Axp(x) + \sum_x bp(x) = A\sum_x xp(x) + b\sum_x p(x)$

$= AE[x] + b$

(d) $\text{cov}(x) = E((x-Ex)(x-Ex)^T)$, Find $\text{cov}(Ax+b)$

Let $Y = Ax+b$, then $\text{cov}(Y) = E[[Y-EY][Y-EY]^T]$

$Y - E[Y] = Ax+b - E[Ax+b] = Ax+b - AE[x] - b = Ax - AE[x] = A(x - E[x])$

$\Rightarrow \text{cov}(Y) = E\left(A(x-E[x])(x-E[x])^TA^T\right) = AE((x-E(x))(x-E(x))^T)A^T = A\,\text{cov}(x)\,A^T$

3. (a) $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $A \in \mathbb{R}^{n\times m}$, find $\nabla_x x^TAy$

Let $z = x^TAy$

$$\frac{\partial z}{\partial x_1} = \frac{\partial \sum_i \sum_j x_i A_{ij} y_j}{\partial x_1} = \frac{\partial(\sum x_1 A_{1j} y_j + \cdots + \sum x_n A_{nj} y_j)}{\partial x_1} = \frac{\partial \sum x_1 A_{1j} y_j}{\partial x_1}$$

$= \sum_j A_{1j} y_j$

$\frac{\partial z}{\partial x} = Ay \qquad \Rightarrow \nabla_x x^TAy = Ay$

(b) $\nabla_y x^TAy$ Let $z = x^TAy$. $\frac{\partial z}{\partial y_1} = \frac{\partial \sum_i \sum_j x_i A_{ij} y_j}{\partial y_1} = \sum_i x_i A_{i1} \Rightarrow \frac{\partial z}{\partial y} = x^TA$

(c) $\nabla_A x^TAy$, Let $z = x^TAy = \sum_i \sum_j x_i A_{ij} y_j$

$\frac{\partial z}{\partial A_{11}} = \frac{\partial \sum_i \sum_j x_i A_{ij} y_j}{\partial A_{11}} = x_1 y_1$

$\frac{\partial z}{\partial A_{12}} = \frac{\partial \sum_i \sum_j x_i A_{ij} y_j}{\partial A_{12}} = x_1 y_2$

$\vdots$

$$\Rightarrow \nabla_A x^TAy = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_m \\ \vdots & & \ddots & \vdots \\ x_ny_1 & & \cdots & x_ny_m \end{bmatrix} = xy^T$$

(d) $f = x^TAx + b^Tx$, find $\nabla_x f$

$\nabla_x f = (A+A^T)x + b$ since $\nabla_x(x^TAx) = \frac{\partial x^TAx}{\partial x} = \frac{\partial \sum a_{ij} x_i x_j}{\partial x}$

$\frac{\partial x^TAx}{\partial x_1} = \sum_{j=1}^n a_{1j}x_j + \sum_{i=1}^n a_{i1}x_i$

$= (Ax)_1 + (A^Tx)_1$

$\Rightarrow \frac{\partial x^TAx}{\partial x} = (A+A^T)x$

for b^Tx part

$b^Tx = b_1x_1 + \cdots + b_nx_n$

$\Rightarrow \frac{\partial b^Tx}{\partial x} = b$

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

(e) $f = tr(AB)$, find $\nabla_A f$

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

$$= tr\begin{bmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_n \\ a_2b_1 & a_2b_2 & & a_2b_n \\ \vdots & & \ddots & \vdots \\ a_nb_1 & \cdots & & a_nb_n \end{bmatrix}$$

$$= \sum_{i=1}^{m} a_{1i}b_{i1} + \sum_{i=1}^{m} a_{2i}b_{i2} + \cdots + \sum_{i=1}^{m} a_{ni}b_{in}$$

$$\Rightarrow \frac{\partial\, tr(AB)}{\partial a_{ij}} = b_{ji}$$

$$\Rightarrow \nabla_A f = \frac{\partial\, tr(AB)}{\partial A} = B^T$$

4. $$L = \min_W \frac{1}{2}\sum_{i=1}^{n} \|y^{(i)} - Wx^{(i)}\|_F^2 = \min_W \frac{1}{2}\sum_{i=1}^{n} (y^{(i)} - Wx^{(i)})^T(y^{(i)} - Wx^{(i)})$$

$$= \min_W Tr\left(\frac{1}{2}(Y - XW)^T(Y - XW)\right)$$

$$\Leftrightarrow \min_W Tr\left[(Y - XW)^T(Y - XW)\right]$$

$$= \min_W Tr(Y^TY - Y^TXW - W^TX^TY + W^TX^TXW)$$

$$= \min_W Tr(Y^TY - 2Y^TXW + W^TX^TXW)$$

$$\nabla_W L = 0 \Rightarrow \frac{\partial Tr(Y^TY - 2Y^TXW + W^TX^TXW)}{\partial W} = 0$$

$$\Rightarrow 0 - 2X^TY + 2X^TXW = 0$$

$$\Rightarrow W = (X^TX)^{-1}X^TY$$ is the optimal W

Scanned by CamScanner