Probability and Statistics
Supervision 3

Samuel Lee

Question 1

(a)

For $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, $\sum_{i=1}^{n} a_i b_i = \mathbf{a}^\intercal \mathbf{b}$.

(b)

For $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^J$, $a_i b_j = [\mathbf{a}\mathbf{b}^\intercal]_{ij}$ where $[\mathbf{a}\mathbf{b}^\intercal]$ is a $n \times J$ matrix.

(c)

For an $n \times J$ matrix $\mathbf{A}$ and $\mathbf{x} \in \mathbb{R}^J$, we have $\sum_{j=1}^{J} a_{ij} x_j = [\mathbf{A}\mathbf{x}]_i$, $i = 1, \ldots, n$ where $[\mathbf{A}\mathbf{x}] \in \mathbb{R}^n$.

(d)

For an $n \times J$ matrix $\mathbf{A}$ and a $J \times K$ matrix $\mathbf{B}$, $\sum_{j=1}^{J} a_{ij} b_{jk} = [\mathbf{AB}]_{ik}$, $i = 1, \ldots, n$; $k = 1, \ldots, K$ where $[\mathbf{AB}]$ is a $n \times K$ matrix.

(e)

For an $n \times J$ matrix $\mathbf{A}$, $K \times J$ matrix $\mathbf{B}$, and $K \times M$ matrix $\mathbf{C}$, $\sum_{j=1}^{j} \sum_{k=1}^{K} a_{ij} b_{kj} c_{km} = [\mathbf{A}\mathbf{B}^\intercal\mathbf{C}]_{im}$, $i = 1, \ldots, n$; $m = 1, \ldots, M$ where $[\mathbf{A}\mathbf{B}^\intercal\mathbf{C}]$ is a $n \times M$ matrix.

(f)

For a $n \times J$ matrix $\mathbf{A}$ and $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^J$, $\sum_{i=1}^{n} \sum_{j=1}^{J} = \mathbf{x}^\intercal \mathbf{A} \mathbf{y}$.

Question 2

If A and B are matrices such that the traces of AB and BA are defined, that is, A is a $n \times m$ matrix and B is a $m \times n$ matrix, then $[AB]_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}$ and $[BA]_{ij} = \sum_{k=1}^{n} b_{ik} a_{kj}$. Therefore the trace of AB is

$$\mathrm{tr}(AB) = \sum_{i=1}^{n} [AB]_{ii} = \sum_{i=1}^{n} \sum_{k=1}^{m} a_{ik} b_{ki}$$

and the trace of BA is

$$\operatorname{tr}(BA) = \sum_{i=1}^{m} [BA]_{ii} = \sum_{i=1}^{m} \sum_{k=1}^{n} b_{ik} a_{ki}$$

which is the same as $\operatorname{tr}(AB)$ in all but notation.

With this result, if the traces of ABC, CAB, and BCA are defined, that is, if A is a $m \times n$ matrix, B is a $n \times k$ matrix, and C is a $k \times m$ matrix, then

$$\operatorname{tr}(ABC) = \operatorname{tr}(XC) = \operatorname{tr}(CX) = \operatorname{tr}(CAB) = \operatorname{tr}(YB) = \operatorname{tr}(BY) = \operatorname{tr}(BCA)$$

where $X = AB$ and $Y = CA$.

Question 3

(a)

$X^\intercal X$ is symmetric since $(X^\intercal X)^\intercal = X^\intercal \times (X^\intercal)^\intercal = X^\intercal X$. Alternatively, we have $[X^\intercal X]_{ij} = [X^\intercal X]_{ji}$, since $[X^\intercal X]_{ij} = \sum_{t=1}^{T} x_{ti} x_{tj}$ if X is a $T \times K$ matrix.

(b)

$X^\intercal X$ is positive semi-definite if, for all $y \in \mathbb{R}^K$, $y^\intercal X^\intercal X y \geq 0$. This is always true since $y^\intercal X^\intercal X y = (Xy)^\intercal X y = \sum_{t=1}^{T} [Xy]_t^2 \geq 0$ where $Xy \in \mathbb{R}^T$.

(c)

$X^\intercal X$ is positive definite if $(Xy)^\intercal X y > 0 \ \forall y \in \mathbb{R}^K$. If X does not have full column rank, then $\exists y$ s.t. $Xy = 0$. The condition for $X^\intercal X$ to be positive definite is for X to have full column rank.

(d)

We have that $(I - X(X^\intercal X)^{-1} X^\intercal) X = 0$. If $I - X(X^\intercal X)^{-1} X^\intercal$ were non-singular, this would mean that $X = 0$, which is a contradiction since X is non-zero. Therefore $I - X(X^\intercal X)^{-1} X^\intercal$ is singular and its determinant is 0.

Question 4

A set of vectors $B = \{b_1, \ldots, b_n\}$ is a basis for some vector space V if $\forall v \in V \ \exists \alpha = \{\alpha_1, \ldots, \alpha_n\}$ s.t. $v = \sum_{i=1}^{n} \alpha_i b_i$, where α is unique. If $\{e_1, e_2, e_3\}$ is a basis for V, $\forall v \in V \ \exists \alpha \in \mathbb{R}^3$ s.t. $v = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$. This means that $\{e_1 + e_2, e_2 + e_3, e_3 + e_1\}$ is also a basis for V since $\forall v \in V$, $v = \beta_1(e_1 + e_2) + \beta_2(e_2 + e_3) + \beta_3(e_3 + e_1)$ where $\beta_1 = \frac{\alpha_1 + \alpha_2 - \alpha_3}{2}$, $\beta_2 = \frac{\alpha_2 + \alpha_3 - \alpha_1}{2}$, and $\beta_3 = \frac{\alpha_3 + \alpha_1 - \alpha_2}{2}$.

Question 5

If $x_1, \ldots, x_K, x$ are linearly dependent, there exists $\alpha_1, \ldots, \alpha_K, \alpha$ such that $\alpha_1 x_1 + \ldots + \alpha_K x_K + \alpha x = 0$, or $\sum_{k=1}^{K} \alpha_k x_k + x = 0$ where we rescale α_k by $\frac{1}{\alpha}$. Therefore, $x = -\sum_{k=1}^{K} \alpha_k x_k$. If this expression were not unique, there would also exist $\alpha_1^*, \dots, \alpha_K^*$ such that $x = -\sum_{k=1}^K \alpha_k^* x_k$, and $\alpha_k - \alpha_k^* \neq 0$ for at least one $k \in \{1, \dots, K\}$. This would imply that

$$x - x = \sum_{k=1}^{K} (\alpha_k - \alpha_k^*) x_k = 0$$

with $\alpha_k - \alpha_k^*$ not all equal to 0. But this is the definition of linear dependence for $x_1, \dots, x_K$. Therefore, this is a contradiction and there can't exist an alternative expression for x if $x_1, \dots, x_K$ are linearly independent.

Question 6

(a)

The eigenvalues of $A = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$ are 1 and 4, the eigenvalue of $B = \begin{pmatrix} 4 & -3 \\ 3 & -2 \end{pmatrix}$ is 1, the eigenvalue of $AB = \begin{pmatrix} 4 & -3 \\ 12 & -8 \end{pmatrix}$ is -2.

(b)

The trace of a matrix is the sum of its eigenvalues. If A and B have positive eigenvalues, then $\mathrm{tr}(A) > 0, \mathrm{tr}(B) > 0$. But $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B) > 0$, and therefore $A + B$ must have at least one positive eigenvalue.

(c)

The eigenvalues of $A = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$ are $3 \pm 2\sqrt{2}$, the eigenvalues of $B = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}$ are $2 \pm \sqrt{2}$, and the eigenvalues of $C = \begin{pmatrix} 1 & 0 \\ 0 & 30 \end{pmatrix}$ are 1 and 30. Since all the eigenvalues are positive, A, B, and C are positive definite. The eigenvalues of $ABC = \begin{pmatrix} 13 & -90 \\ 5 & -30 \end{pmatrix}$ are -5 and -12.

Question 7

(a)

$$(I_n - M)^2 = (I_n - M)(I_n - M) = I_n(I_n - M) - M(I_n - M) = I_n - M - M + M^2 = I_n - M$$

(b)

$$(I_n - M)(I_n - tM) = I_n(I_n - M) - tM(I_n - M) = I_n - M - tM + tM^2 = I_n - M$$

(c)

$$(2M - I_n)^2 = 2M(2M - I_n) - I_n(2M - I_n) = 4M^2 - 2M - 2M + I_n = I_n$$

(d)

It seems sensible to guess that the inverse of $I_n + M$ will take the form of $\alpha I_n + \beta M$. If so,

$$(I_n + M)(\alpha I_n + \beta M) = \alpha I_n + \beta M + M(\alpha I_n + \beta M) = \alpha I_n + (\alpha + 2\beta)M$$

which is equal to I_n if $\alpha = 1$ and $\beta = \frac{1}{2}$. Indeed,

$$(I_n + M)\left(I_n - \frac{1}{2}M\right) = I_n - \frac{1}{2}M + M\left(I_n - \frac{1}{2}M\right) = I_n$$

Hence the inverse of $I_n + M$ is $I_n - \frac{1}{2}M$.

(e)

With $M = U\Lambda U^{-1}$ and $M = M^2$,

$$M = M^2 \implies U\Lambda U^{-1}U\Lambda U^{-1} = U\Lambda^2 U^{-1} = U\Lambda U^{-1} \implies \Lambda = \Lambda^2$$

Which can only be true if the elements of Λ are 0 or 1.

Question 8

If u is an eigenvector of X with associated eigenvalue λ, then $Xu = \lambda u \implies X^2 u = \lambda X u = \lambda^2 u$. Therefore,

$$X = U\Lambda U^{-1} \implies X^2 = U\Lambda^2 U^{-1}$$

and the eigenvectors of X^2 are the same as the eigenvectors of X, whereas the associated eigenvalues in X^2 are the squares of their counterparts in X.

If $2X^2 - 3X + I_n = 0$, then

$$2U\Lambda^2U^{-1} - 3U\Lambda U^{-1} + I_n = 0_n \implies U \begin{pmatrix} 2\lambda_1^2 - 3\lambda_1 & 0 & \cdots & 0 \\ 0 & 2\lambda_2^2 - 3\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2\lambda_n^2 - 3\lambda_n \end{pmatrix} U^{-1} + I_n = 0_n$$

which means that any matrix X with only eigenvalues such that $2\lambda_i^2 - 3\lambda_i = -1$ for $i \in \{1, \ldots, n\}$ is a solution, since this would mean

$$2X^2 - 3X + I_n = U \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix} U^{-1} + I_n = -UI_nU^{-1} + I_n = -UU^{-1} + I_n = 0_n$$

The eigenvalues which satisfy this are $\lambda = \frac{1}{2},\ \lambda = 1$. This trivially includes $\frac{1}{2}I_n$ and I_n.