LINEAR MODEL NOTES

CARSON JAMES

Contents

1. Matrix Algebra

1.1. Column and Null Space.

Exercise 1.1. *Let $X \in \mathcal{M}_{m,n}$. Then $\mathcal{N}(X) = \mathcal{N}(X^T X)$.*

Proof. Let $a \in \mathcal{N}(X)$. Then $Xa = 0$. So $X^T Xa = 0$. Thus $a \in \mathcal{N}(X^T X)$. Conversely, suppose that $a \in \mathcal{N}(X^T X)$. Then $X^T Xa = 0$. So

$$\begin{aligned} 0 &= a^T X^T Xa \\ &= (Xa)^T (Xa) \\ &= \|Xa\|^2 \end{aligned}$$

Hence $Xa = 0$ and $a \in \mathcal{N}(X)$. □

Exercise 1.2. *Let $X \in \mathcal{M}_{m,n}$. Then $\mathcal{C}(X^T) = \mathcal{C}(X^T X)$.*

Proof.

$$\begin{aligned} \mathcal{C}(X^T) &= \mathcal{N}(X)^\perp \\ &= \mathcal{N}(X^T X)^\perp \\ &= \mathcal{C}(X^T X) \end{aligned}$$

□

Exercise 1.3. *Let $X \in \mathcal{M}_{m,n}$. If $X^T X = 0$, then $X = 0$.*

Proof. Suppose that $X^T X = 0$. Then

$$\begin{aligned} rank(X^T) &= \dim \mathcal{C}(X^T) \\ &= \dim \mathcal{C}(X^T X) \\ &= rank(X^T X) \\ &= 0 \end{aligned}$$

So $X^T = X = 0$. □

Exercise 1.4. *Let $X \in \mathcal{M}_{m,n}$ and $A, B \in \mathcal{M}_{n,p}$. Then $X^T XA = X^T XB$ iff $XA = XB$.*

Proof. Clearly if $XA = XB$, then $X^T XA = X^T XB$. Conversely, suppose that $X^T XA = X^T XB$. Then $X^T X(A - B) = 0$. So for each $i = 1, \cdots, p$, $X^T X(A - B)e_i = 0$. Thus for each $i = 1, \cdots, p$ $X(A - B)e_i \in \mathcal{N}(X^T) \cap \mathcal{C}(X) = \{0\}$. Hence $X(A - B) = 0$ and $XA = XB$. □

1.2. **Generalized Inverses.**

Definition 1.5. *Let $A \in \mathcal{M}_{m,n}$ and $G \in \mathcal{M}_{n,m}$. Then G is said to be a* ***generalized inverse*** *of A if $AGA = A$.*

Theorem 1.6. *Let $A \in \mathcal{M}_{m,n}$. Suppose that $rank(A) = r$. Then there exists $P \in \mathcal{M}_{m,m}, Q \in \mathcal{M}_{n,n}, C \in \mathcal{M}_{r,r}$ such that P, Q, C are non-singular, $rank(C) = r$ and*

$$A = P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q$$

Exercise 1.7. *Let*

$$A = P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q$$

as in the previous theorem and $D \in \mathcal{M}_{r,m-r}, E \in \mathcal{M}_{n-r,r}, F \in \mathcal{M}_{n-r,m-r}$. Put

$$G = Q^{-1} \begin{pmatrix} C^{-1} & D \\ E & F \end{pmatrix} P^{-1}$$

Then G is a generalized inverse of A.

Proof.

$$\begin{aligned} AGA &= \left[P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q \right] \left[Q^{-1} \begin{pmatrix} C^{-1} & D \\ E & F \end{pmatrix} P^{-1} \right] \left[P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q \right] \\ &= P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} C^{-1} & D \\ E & F \end{pmatrix} \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q \\ &= P \begin{pmatrix} I & CD \\ 0 & 0 \end{pmatrix} \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q \\ &= P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q \\ &= A \end{aligned}$$

□

Note 1.8. *The previous exercise and theorem guarantee the existence of a generealized inverse for all matrices. We will denote a generalized inverse of A by A^-.*

Theorem 1.9. *Let $A \in \mathcal{M}_{m,n}$. Suppose that $rank(A) = r$. Let $P \in \mathcal{M}_{mm}$, $Q \in \mathcal{M}_{n,n}$ permutation matrices and $C \in \mathcal{M}_{r,r}$. Suppose that $rank(C) = r$ and $PAQ = \begin{pmatrix} C & D \\ E & F \end{pmatrix}$ Then $Q \begin{pmatrix} C^{-1} & 0 \\ 0 & 0 \end{pmatrix} P$ is a generalized inverse of A.*

Exercise 1.10. *Let $X \in \mathcal{M}_{m,n}$. Then $(X^T)^- = (X^-)^T$.*

Proof.

$$\begin{aligned} X^T(X^-)^T X^T &= (XX^-X)^T \\ &= X^T \end{aligned}$$

□

Exercise 1.11. *Let $X \in \mathcal{M}_{m,n}$. Then $\mathcal{C}(XX^-) = \mathcal{C}(X)$.*

Proof. Clearly $\mathcal{C}(XX^-) \subset \mathcal{C}(X)$. Let $b \in \mathcal{C}(X)$. Then there exists $a \in \mathbb{R}^n$ such that $Xa = b$. Then

$$\begin{aligned} XX^-b &= XX^-Xa \\ &= Xa \\ &= b \end{aligned}$$

So $b \in \mathcal{C}(XX^-)$. Thus $\mathcal{C}(X) \subset \mathcal{C}(XX^-)$ and $\mathcal{C}(X) = \mathcal{C}(XX^-)$ □

Exercise 1.12. *Let $X \in \mathcal{M}_{m,n}$. Then $\mathcal{N}(X) = \mathcal{N}(X^-X)$*

Proof. From the previous exercise, we have that

$$\begin{aligned} \mathcal{N}(X) &= \mathcal{C}(X^T)^\perp \\ &= \mathcal{C}(X^T(X^T)^-)^\perp \\ &= \mathcal{C}(X^T(X^-)^T)^\perp \\ &= \mathcal{C}((X^-X)^T)^\perp \\ &= \mathcal{N}(X^-X) \end{aligned}$$

□

Exercise 1.13. *Let $X \in \mathcal{M}_{m,n}$. Then $X^- = (X^TX)^-X^T$.*

Proof. By definition, $X^TX(X^TX)^-X^TX = X^TX$. A previous exercise implies that $X(X^TX)^-X^TX = X$. Thus $X^- = (X^TX)^-X^T$. □

Exercise 1.14. *Let $X \in \mathcal{M}_{m,n}$. Then $(X^T)^- = X(X^TX)^-$.*

Proof. The previous exercise tells us that $X^- = (X^TX)^-X^T$. Transposing both sides, we obtain $(X^T)^- = X(X^TX)^-$. □

1.3. **Projections.**

Definition 1.15. *Let $A \in \mathcal{M}_{m,m}$. Then X is said to be **idempotent** if $A^2 = A$.*

Exercise 1.16. *Let $X \in \mathcal{M}_{m.n}$. Then XX^- and X^-X are idempotent*

Proof.

$$\begin{aligned}(XX^-)(XX^-) &= (XX^-X)X^- \\ &= XX^-\end{aligned}$$

The case is similar for X^-X. □

Exercise 1.17. *Let $A \in \mathcal{M}_{m.m}$. If X is idempotent, then $I - A$ is idempotent.*

Proof. Suppose that A is idempotent. Then

$$\begin{aligned}(I - A)(I - A) &= I^2 - IA - AI + A^2 \\ &= I - 2A + A \\ &= I - A\end{aligned}$$

□

Theorem 1.18. *Let $A \in \mathcal{M}_{m,m}$. If A is idempotent, then $rank(A) = tr(A)$.*

Definition 1.19. *Let $P \in \mathcal{M}_{m,m}$ and $S \subset \mathbb{R}^m$ a subspace. Then P is said to be a **projection matrix** onto S if*

(1) P is idempotent
(2) $\mathcal{C}(X) \subset S$
(3) for each $x \in S$, $Px = x$

Note 1.20. *In the previous definition, (2) and (3) imply that $\mathcal{C}(X) = S$, so to say that X projects "onto" S is accurate.*

Exercise 1.21. *Let $S \subset \mathbb{R}^m$ and P, Q projection matrices onto S. Then $PQ = Q$.*

Proof. Let $x \in \mathbb{R}^m$. Then $Qx \in \mathcal{C}(Q) = S$. So $PQx = Qx$. Thus $PQ = Q$. □

Exercise 1.22. *Let $X \in \mathcal{M}_{m,n}$. Then XX^- is a projection onto $\mathcal{C}(X)$.*

Proof. A previous exercises tells us that XX^- is idempotent. Another previous exercise tells us that $\mathcal{C}(XX^-) = \mathcal{C}(X)$. Let $b \in \mathcal{C}(X)$. Then there exists $a \in \mathbb{R}^n$ such that $Xa = b$. So

$$\begin{aligned}XX^-b &= XX^-Xa \\ &= Xa \\ &= b\end{aligned}$$

□

Exercise 1.23. *Let $X \in \mathcal{M}_{m,n}$. Then $I - X^-X$ is a projection onto $\mathcal{N}(X)$*

Proof. Since X^-X is idempotent, so is $I - X^-X$. Let $b \in \mathcal{C}(I - X^-X)$. Then there exists $a \in \mathbb{R}^n$ such that $(I - X^-X)a = b$. Then

$$\begin{aligned} Xb &= X(I - X^-X)a \\ &= (X - XX^-X)a \\ &= (X - X)a \\ &= 0a \\ &= 0 \end{aligned}$$

So $\mathcal{C}(I - X^-X) \subset \mathcal{N}(X)$. Let $a \in \mathcal{N}(X)$. Then $Xa = 0$ and

$$\begin{aligned} (I - X^-X)a &= a - X^-Xa \\ &= a \end{aligned}$$

So for each $a \in \mathcal{N}(X)$, $(I - X^-X)a = a$. □

Exercise 1.24. *Let $S \subset \mathbb{R}^m$ be a subspace and $P \in \mathcal{M}_{m,m}$ be a symmetric projection matrix onto S. Then P is unique.*

Proof. Let $Q \in \mathcal{M}_{m,m}$ be a symmetric projection matrix onto S. Then

$$\begin{aligned} (P - Q)^T(P - Q) &= P^TP - P^TQ - Q^TP + Q^TQ \\ &= P^2 - PQ - QP + Q^2 \\ &= P - Q - P + Q \\ &= 0 \end{aligned}$$

Thus $P - Q = 0$ and $P = Q$. □

Definition 1.25. *Let $X \in \mathcal{M}_{m,n}$. We define P_X by*

$$P_X = X(X^TX)^-X^T$$

Exercise 1.26. *Let $X \in \mathcal{M}_{m,n}$. Then P_X is well defined. That is, independent of the choice of $(X^TX)^-$.*

Proof. Suppose that G, H are generalized inverses of X^TX. By definition, we have

$$\begin{aligned} X^TXGX^TX = X^TXHX^TX &\Rightarrow XGX^TX = XHX^TX \\ &\Rightarrow X^TXG^TX^T = X^TXHX^T \\ &\Rightarrow XG^TX^T = XHX^T \\ &\Rightarrow XGX^T = XHX^T = P_X \end{aligned}$$

□

Note 1.27. *Recall that $X^- = (X^TX)^-X^T$. So that $P_X = XX^-$ is indeed a projection onto $\mathcal{C}(X)$. Recall that $[(X^TX)^-]^T$ is a gen. inv. of $(X^TX)^T = (X^TX)$. Hence $P_X^T = X[(X^TX)^-]^TX^T = P_X$. Since P_X is symmetric, it is the unique symmetric projection onto $\mathcal{C}(X)$.*

Note 1.28. *Recall that $(X^T)^- = X(X^TX)^-$. So that $P_X = (X^T)^-X^T$. A previous exercises tells us that $I - P_X$ is a projection on $\mathcal{N}(X^T)$. Since $I - P_X$ is symmetric, it is the unique symmetric projection onto $\mathcal{N}(X^T)$.*

1.4. **Solving Linear Equations.**

Definition 1.29. *Let $A \in \mathcal{M}_{m,n}$ and $b \in \mathbb{R}^m$. Then the system $Ax = b$ is said to be* ***consistent*** *if $b \in \mathcal{C}(A)$.*

Exercise 1.30. *Let $A \in \mathcal{M}_{m,n}$ and $G \in \mathcal{M}_{n,m}$. Then $G = A^-$ iff for each $b \in \mathcal{C}(A)$, Gb solves $Ax = b$.*

Proof. Suppose that $G = A^-$. Let $b \in \mathcal{C}(A)$. Then there exists $x^* \in \mathbb{R}^n$ such that $Ax^* = b$. So

$$\begin{aligned} A(Gb) &= AG(Ax^*) \\ &= (AGA)x^* \\ &= Ax^* \\ &= b \end{aligned}$$

So Gb solves $Ax = b$. Conversely, Suppose that for each $b \in \mathcal{C}(A)$, Gb solves $Ax = b$. Let $z \in \mathbb{R}^n$. So $Az \in \mathcal{C}(A)$. Then

$$\begin{aligned} (AGA)z &= A[G(Az)] \\ &= Az \end{aligned}$$

Since for each $z \in \mathbb{R}^n AGAz = Az$, $AGA = A$ and $G = A^-$. □

Exercise 1.31. *Let $b \in \mathcal{C}(A)$. Then*

$$\{x \in \mathbb{R}^n : Ax = b\} = \{A^- b + (I - A^- A)z : z \in \mathbb{R}^n\}$$

.

Proof. Let $x \in \{A^- b + (I - A^- A)z : z \in \mathbb{R}^n\}$. Then there exists $z \in \mathbb{R}^n$ such that $x = A^- b + (I - A^- A)z$. Since $(I - A^- A)$ is a projection onto $\mathcal{N}(A)$,

$$\begin{aligned} Ax &= AA^- b \\ &= b \end{aligned}$$

So $x \in \{x \in \mathbb{R}^n : Ax = b\}$. Conversely, let $x \in \{x \in \mathbb{R}^n : Ax = b\}$. Then

$$\begin{aligned} x &= A^-(Ax) + (x - A^- Ax) \\ &= A^-(b) + (I - A^- A)x \\ &\in \{A^- b + (I - A^- A)z : z \in \mathbb{R}^n\} \end{aligned}$$

□

1.5. **Moore-Penrose Pseudoinverse.**

Theorem 1.32. *(Singular Value Decomposition):*
Let $A \in \mathcal{M}_{m,n}$. Suppose that $rank(A) = r$. Then there exist $U \in \mathcal{M}_{m,m} V \in \mathcal{M}_{n,n}$, and $D_0 \in \mathcal{M}_{r,r}$ such that

(1) $A = U \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} V^T$
(2) $U^T U = I$
(3) $V^T V = I$

(4) $D_0 = diagonal(d_1, d_2, \cdots, d_r)$ *with* $d_1 \geq d_2 \geq \cdots \geq d_r > 0$

Note 1.33. *Put* $D = \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{m,n}$

(1) Since D_0 *is symmetric,* $D^T = \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{n,m}$

(2) Since D_0 *is diagonal,* D_0^{-1} *is also diagonal and symmetric*

Definition 1.34. *Let* $A \in \mathcal{M}_{m,m}$ *and* $A^+ \in \mathcal{M}_{n,m}$*. Then* A^+ *is said to be a* ***Moore-Penrose pseudoinverse*** *of* A *if*

(1) $AA^+A = A$
(2) $A^+AA^+ = A^+$
(3) AA^+ *is symmetric*
(4) A^+A *is symmetric*

Note 1.35. *We have that* $P_X = XX^+ = X(X^TX)^-X^T$*.*

Exercise 1.36. *Let* $A \in \mathcal{M}_{m,n}$ *and* $S, T \in \mathcal{M}_{n,m}$*. If* S *and* T *are m-p pseudoinverses of* A*, then* $S = T$*.*

Proof. Suppose that S, T satisfy properties (1)-(4). Then

$$\begin{aligned} S &= SAS \\ &= (SA)^TS \\ &= A^TS^TS \\ &= (ATA)^TS^TS \\ &= A^TT^TA^TS^TS \\ &= (TA)^T(SA)^TS \\ &= (TA)(SA)S \\ &= TA(SAS) \\ &= TAS \end{aligned}$$

and

$$\begin{aligned} T &= TAT \\ &= T(AT)^T \\ &= TT^TA^T \\ &= TT^T(ASA)^T \\ &= TT^TA^TS^TA^T \\ &= T(AT)^T(AS)^T \\ &= T(AT)(AS) \\ &= (TAT)AS \\ &= TSA \end{aligned}$$

So $S = T$ ∎

Exercise 1.37. *Let $A \in \mathcal{M}_{m,n}$ have singular value decomposition $A = UDV^T = U \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} V^T$. Define $D^+ = \begin{pmatrix} D_0^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{n,m}$. Then D^+ is the m-p pseudoinverse of D.*

Proof.

(1)

$$
\begin{aligned}
DD^+D &= \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} D_0^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \\
&= D
\end{aligned}
$$

(2) Similar to (1).

(3)

$$
\begin{aligned}
(DD^+)^T &= \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}^T \\
&= \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \\
&= DD^+
\end{aligned}
$$

(4) Similar to (3).

□

Exercise 1.38. *Let $A \in \mathcal{M}_{m,n}$ have singular value decomposition $A = UDV^T$. So $A^T \in \mathcal{M}_{n,m}$ has singular value decomposition $A^T = VD^TU^T$. Then $(D^T)^+ = (D^+)^T$*

Proof. Since $D^T = \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{n,m}$, we have that $(D^T)^+ = \begin{pmatrix} D_0^{-1} & 0 \\ 0 & 0 \end{pmatrix} = (D^+)^T$ □

Exercise 1.39. *Let $A \in \mathcal{M}_{m,n}$ have singular value decomposition $A = UDV^T$. Define $A^+ = VD^+U^T$. Then A^+ is the m-p pseudoinverse of A.*

Proof. (1)

$$
\begin{aligned}
AA^+A &= (UDV^T)(VD^+U^T)(UDV^T) \\
&= UDD^+DV^T \\
&= UDV^T \\
&= A
\end{aligned}
$$

(2) Similar to (1)

(3)

$$
\begin{aligned}
(AA^+)^T &= \left[(UDV^T)(VD^+U^T)\right]^T \\
&= (UDD^+U^T)^T \\
&= U(DD^+)^T U^T \\
&= UDD^+U^T \\
&= (UDV^T)(VD^+U^T) \\
&= AA^+
\end{aligned}
$$

(4) Similar to (3).

□

Exercise 1.40. *Let $A \in \mathcal{M}_{m,n}$ have singular value decomposition $A = UDV^T$. Then $(A^T)^+ = (A^+)^T$.*

Proof.

$$
\begin{aligned}
(A^T)^+ &= \left[(UDV^T)^T\right]^+ \\
&= (VD^TU^T)^+ \\
&= U(D^T)^+V^T \\
&= U(D^+)^TV^T \\
&= (VD^+U^T)^T \\
&= (A^+)^T
\end{aligned}
$$

□

Exercise 1.41. *Let $A \in \mathcal{M}_{m,n}$. Then there exists a unique matrix $A^+ \in \mathcal{M}_{n,m}$ such that A^+ is the m-p pseudoinverse of A.*

Proof. The existence of and uniqueness of A^+ are shown in the previous exercises. □

Exercise 1.42. *Let $A \in \mathcal{M}_{m,m}$. Then $(A^+)^+ = A$.*

Proof. We observe that A satisfies properties $(1)-(4)$ for A^+. By uniqueness, $(A^+)^+ = A$. □

Exercise 1.43. *Let $A \in \mathcal{M}_{m,n}$ and $b \in \mathcal{C}(A)$. Pur $S = \{x \in \mathbb{R}^n : Ax = b\}$. Then*

$$\|A^+b\| = \min_{x \in S} \|x\|$$

.

Proof. Let $x \in S$. A previous exercise tells us that there exists $z \in \mathbb{R}^n$ such that $x = A^+b + (I - A^+A)z$. Then

$$\begin{aligned}
\|x\|^2 &= \|A^+b + (I - A^+A)z\|^2 \\
&= (A^+b + (I - A^+A)z)^T(A^+b + (I - A^+A)z) \\
&= \|A^+b\|^2 - 2z^T(I - A^+A)^T(A^+b) + \|(I - A^+A)z\|^2 \\
&= \|A^+b\|^2 - 2z^T(I - A^+A)A^+b + \|(I - A^+A)z\|^2 \\
&= \|A^+b\|^2 + \|(I - A^+A)z\|^2 \\
&\geq \|A^+b\|^2
\end{aligned}$$

□

1.6. **Differentiation.**

Definition 1.44. *Let $Q : \mathbb{R}^n \to \mathbb{R}$ given by $b \mapsto Q(b)$. Suppose that $Q \in C^1(\mathbb{R}^n)$. We define*

$$\frac{\partial Q}{\partial b} = \begin{pmatrix} \frac{\partial Q}{\partial b_1} \\ \vdots \\ \frac{\partial Q}{\partial b_n} \end{pmatrix}$$

Exercise 1.45. *Let $a, b \in \mathbb{R}_n$ and $A \in \mathcal{M}_{n,n}$. Then*

(1)
$$\frac{\partial a^T b}{\partial b} = a$$

(2)
$$\frac{\partial b^T A b}{\partial b} = (A + A^T)b$$

Proof.

(1) Since
$$a^T b = \sum_{i=1}^n a_i b_i$$
We have that
$$\frac{\partial a^T b}{\partial b_i} = a_i$$
and therefore
$$\frac{\partial a^T b}{\partial b} = a$$

(2) Since
$$\begin{aligned}
b^T A b &= \sum_{i=1}^n b_i \sum_{j=1}^n A_{i,j} b_j \\
&= \sum_{i=1}^n \sum_{j=1}^n b_i A_{i,j} b_j
\end{aligned}$$

The terms containing b_i are

$$A_{i,i}b_i^2 + \sum_{\substack{j=1 \\ j\neq i}}^{n}(A_{i,j} + A_{j,i})b_i b_j$$

This implies that

$$\begin{aligned}\frac{\partial b^T Ab}{\partial b_i} &= 2A_{i,i}b_i + \sum_{\substack{j=1 \\ j\neq i}}^{n}(A_{i,j} + A_{j,i})b_j \\ &= \sum_{j=1}^{n}(A_{i,j} + A^T_{i,j})b_j \\ &= [(A + A^T)b]_i\end{aligned}$$

So

$$\frac{\partial b^T Ab}{\partial b} = (A + A^T)b$$

$\square$

2. The Linear Model

2.1. Model Description.

Definition 2.1. *Given $y \in \mathbb{R}_m$ a vector of observed responses to the matrix $X \in \mathcal{M}_{m,n}$ of observed inputs, we will consider the model*

$$y = Xb + e$$

where $b \in \mathbb{R}_n$ is a vector of unknown parameters and $e \in \mathbb{R}^m$ is a random vector of unobserved errors with zero mean.

Definition 2.2. *For a parameter vector $b \in \mathbb{R}^n$, we have that $e = y - Xb$. For this reason, e is called the* ***residual vector*** *or simply the "residuals".*

Note 2.3. *The goal will be to find a parameter vector $b \in \mathbb{R}^n$ that makes the causes the residuals to be as small as possible.*

2.2. Least Squares Optimization.

Definition 2.4. *We define the* ***cost function****, $Q : \mathbb{R}^n \to \mathbb{R}$ by*

$$\begin{aligned}Q(b) &= \|y - Xb\|^2 \\ &= (y - Xb)^T(y - Xb)\end{aligned}$$

Definition 2.5. *Let $b \in \mathbb{R}^n$. Then b is said to be a* ***least squares solution*** *for the model if*

$$Q(b) = \inf_{c\in R^n} Q(c)$$

Exercise 2.6. *If b is a least squares solution for the model, then $X^TXb = X^Ty$.*

Proof. Suppose that b is a least squares solution for the model, then Q has a local minimum at b. Since Q is convex in b, this global minimum is also a local minimum. Thus

$$\frac{\partial Q}{\partial b}(b) = 0$$

By definition,

$$\begin{aligned} Q(b) &= y^T y - y^T Xb - b^T X^T y + b^T X^T Xb \\ &= y^T y - 2y^T Xb + b^T X^T Xb \end{aligned}$$

Thus

$$\begin{aligned} 0 &= \frac{\partial Q}{\partial b}(b) \\ &= -2X^T y + 2X^T Xb \end{aligned}$$

Hence $X^T Xb = X^T y$. □

Definition 2.7. *For $y \in \mathbb{R}^m$ and $X \in \mathcal{M}_{m,n}$, we define the* ***normal equation*** *to be*

$$X^T Xb = X^T y$$

Exercise 2.8. *The normal equation is consistent.*

Proof. We have that $X^T y \in \mathcal{C}(X^T) = \mathcal{C}(X^T X)$. □

Exercise 2.9. *Let $b \in \mathbb{R}^n$. Then b is a least squares solution for the model iff b satisfies the normal equation.*

Proof. The previous exercises tells us that if b is a least squares solution for the model, then b satisfies the normal equation. Conversely, suppose that b satisfies the normal equation. Then

$$\begin{aligned} Q(c) &= (y - Xc)^T (y - Xc) \\ &= (y - Xb + Xb - Xc)^T (y - Xb + Xb - Xc) \\ &= (y - Xb)^T (y - Xb) - (y - Xb)^T (X(b - c)) - (b - c)^T X^T (y - Xb) + (b - c)^T X^T (X(b - c)) \\ &= Q(b) - 2(b - c)^T X^T (y - Xb) + \|X(b - c)\|^2 \\ &= Q(b) + \|X(b - c)\|^2 \end{aligned}$$

Thus b minimizes Q. □

Exercise 2.10. *Let $b \in \mathbb{R}^n$ be a least squares solution for the model. Then $\|y\|^2 = \|Xb\|^2 + \|e\|^2$*

Proof. Since b satisfies the normal equation, we have that $X^T(y - Xb) = 0$. Thus

$$\begin{aligned} Xb \cdot e &= b^T X^T e \\ &= b^T X^T (y - Xb) \\ &= b^T 0 \\ &= 0 \end{aligned}$$

So Xb and e are orthogonal. Therefore

$$\begin{aligned} \|y\|^2 &= \|Xb + e\|^2 \\ &= \|Xb\|^2 + \|e\|^2 \end{aligned}$$

□