SLMR Errata

Joe Suzuki

June 8, 2024

1 Linear Algebra

1. P5 L2-:
$$\begin{bmatrix} 0 & a_2 - a_1 & \dots & a_2^{k-2} - a_1^{k-2} & a_2^{k-1} - a_1^{k-1} \end{bmatrix}$$
2. P5 L1-:
$$\begin{vmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & a_2 - a_1 & \dots & (a_2 - a_1)\, a_2^{k-3} & (a_2 - a_1)\, a_2^{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & a_k - a_1 & \dots & (a_k - a_1)\, a_k^{k-3} & (a_k - a_1)\, a_2^{k-2} \end{vmatrix}$$
3. P8 Footnote: (1.2) $\mathbb{R}$ is said $\to$ (1.2) is said
4. P9 Example 13: the matrix A in Example 8 $\to$ the matrix A below
5. P12 L12: Should be $u_1 = \frac{1}{\|v_1\|} v_1 \to u_1 = \frac{v_1}{\|v_1\|}$
6. P13 L9: symmetric because $\to$ symmetric because when A is symmetric
7. P13 Example 18: The eigenspaces of $\to$ The eigenspaces of $A =$
8. P15 L10: $b_{r+1}, \dots, b_n \to b_1, \dots, b_n$
9. P15 Proposition 7: $P^{-}1 \to P^{-1}$
10. P15 L17: Proposition (1.6) $\to$ Proposition 4
11. P15 L7-: For a nonsingular matrix $\to$ For a matrix

2 LINEAR REGRESION

1. P21 L3: `p=2` should be inserted.
2. P23 L1-: $\sigma_{i,j} := E(\hat\beta_i - \beta_i)(\hat\beta_j - \beta_j)^\top \to \sigma_{i,j} := E(\hat\beta_i - \beta_i)(\hat\beta_j - \beta_j)$
3. P24 L5,L6:
$$= E(\hat\beta - \beta)(\hat\beta - \beta)^T = E\left\{\left(X^T X\right)^{-1} X^T \epsilon\right\}\left\{\left(X^T X\right)^{-1} X^T \epsilon\right\}^T$$
$$= \left(X^T X\right)^{-1} X^T \left(E\epsilon\epsilon^T\right) X \left(X^T X\right)^{-1} = \sigma^2 \left(X^T X\right)^{-1}$$

4. P25 L10-: $[v_1,\ldots,v_{N-p-1},v_{N-p},\ldots,v_n] \to [v_1,\ldots,v_{N-p-1},v_{N-p},\ldots,v_N]$
5. P26 Section Title: $\hat{\beta}_j = 0 \to \beta_j = 0$
6. P39 L10, L12: Proposition 4 $\to$ Proposition 3

3 ClASSIFICATION

1. P48 L1-: $x > -\beta_0/\beta \to x > -\beta_0/\beta$ for $\beta > 0$
2. P52 L2: below: $\mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^p \times \{-1,1\}$
3. P54 L6: $\det \Sigma \to \Sigma_\pm$
4. P54 L6-: $1-\sum_{k\neq\hat{k}} P(y=k|x) = 1-P(y=\hat{k}|x) \to 1-\sum_{k\neq\hat{k}} P(y=k|x) = P(y=\hat{k}|x)$
5. P55 L6: $\mu_{-1}\Sigma^{-1}\mu_{-1} \to \mu_{-1}^\top\Sigma^{-1}\mu_{-1}$
6. P55 L9: $\pi_1 = \pi_{-1} \to \pi_1 = \pi_{-1}, p = 1$
7. P55 Program L9: `mean(c(x.1,y.1))` $\to$ `c(mean(x.1),mean(y.1))`,
   `mean(c(x.2,y.2))` $\to$ `c(mean(x.2),mean(y.2))`
8. P55 L2-: `mat =cov(df); inv.1= solve(mat);de.1=det(mat); inv.2=inv.1; de.2=de.1`
9. P56 L11- and P64 Exercise 29: `z=array(dim=n/2)` $\to$ `z=array(dim=length(test))`
10. P59 L7-,L8- (four) and P65 Exercise 30 (two): `pnorm` $\to$ `dnorm`.
11. P57 L5: $\{1,\ldots,n\} \to \{1,\ldots,N\}$
12. P57 L11: $S=\{2\} \to S=\{1\}$.

4 RESAMPLING

1. P69 L6: `cv.linear(X[,c(1,4,5)],y,10)` $\to$ `cv.linear(X[,c(1,4,5,6)],y,10)`
2. P70 L5-: Chap. 2 $\to$ Chap. 3
3. P71 L4- (in the Program):
   `knn.ans = knn(df[-index,1:4],df[index, 1:4],df[-index, 5], k)`
   Remove library(class)
4. P71 L3-: $i \notin S \to i \in S$
5. P74 L1: Section 2.6 $\to$ Section 2.7.
6. P74 L3 below: $\hat{\sigma}^2(\hat{\alpha}) := \dfrac{1}{r-1}\displaystyle\sum_{h=1}^r \{\hat{\alpha}_h - \hat{\alpha}\}^2 \to \hat{\sigma}^2(\hat{\alpha}) := \dfrac{1}{r-1}\displaystyle\sum_{h=1}^r \{\hat{\alpha}_h - \dfrac{1}{r}\sum_{j=1}^r \hat{\alpha}_j\}^2$
7. P75 3-: $\alpha = \frac{V(Y)-v(Y)}{V(X)+V(Y)-2Cov(X,Y)} \to \alpha = \dfrac{V(Y)-Cov(X,Y)}{V(X)+V(Y)-2Cov(X,Y)}$

8. P76 2: $\hat{\alpha} = \dfrac{v_x^2 - v_y^2}{v_x^2 + v_y^2 - 2c_{x,y}} \to \hat{\alpha} = \dfrac{v_y^2 - c_{xy}}{v_x^2 + v_y^2 - 2c_{x,y}}$

9. P76 7: `return((V(Y)-v(X))/(V(X)+V(Y)-2Cov(X,Y)))` $\to$ `return((V(Y)-Cov(X,Y))/(V(X)+V(Y)-2Cov(X,Y)))`

10. P78 L5: Proposition 17 $\to$ Proposition 15

11. P78 L14 Proposition 18 $\to$ Proposition 16

12. P79 Exercise 33(b): $\hat{y}_S = X_S\hat{\beta}_S \to \hat{y}_S = X_S\hat{\beta}_{-S}$

5 Information Criteria

1. P85 L20-: `q=T[,j]; S=sum((lm(y~X[,q])$fitted.values-y)^2)/n` $\to$ `q=T[,j]; S=sum((lm(y~X[,q])$fitted.values-y)^2)`

2. P86 L2, L3: `S.min` $\to$ `res$value/n` (two)

3. P86 L3: `log N` $\to$ `log n`

4. L86 L8: `TSS/(n-1)` $\to$ `TSS*(n-1)`

5. P87 L1-, Exercise P97 L3: $f(y|x,\beta) := \dfrac{1}{\sqrt{(2\pi)^{p/2}}}\exp\{-\dfrac{1}{2\sigma^2}\|y - x\beta\|\}$

   $\to f(y_i|x_i,\beta) := \dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\dfrac{1}{2\sigma^2}(y_i - x_i\beta)^2\}$

6. P88 L9, Exercise 40(b): $\dfrac{\partial l^2}{\partial \sigma^2} = -\dfrac{N}{2\sigma^2} - \dfrac{\|y - X\beta\|^2}{2(\sigma^2)^2} \to \dfrac{\partial l}{\partial \sigma^2} = -\dfrac{N}{2\sigma^2} + \dfrac{\|y - X\beta\|^2}{2(\sigma^2)^2}$

7. P89 (5.3)(5.4) (six): $x \to X$

8. P89 Example 47: $\nabla l = -\dfrac{1}{\sigma^2}\sum_{i=1}^N x_i^T(y_i - x_i\beta) \to \nabla l = \dfrac{1}{\sigma^2}\sum_{i=1}^N x_i^T(y_i - x_i\beta)$,

   $\nabla^2 l = \dfrac{1}{\sigma^2}\sum_{i=1}^N x_i^T x_i = \dfrac{1}{\sigma^2}X^TX \to \nabla^2 l = -\dfrac{1}{\sigma^2}\sum_{i=1}^N x_i^T x_i = -\dfrac{1}{\sigma^2}X^TX$,

   $E[\nabla l] = -\dfrac{1}{\sigma^2}\sum_{i=1}^N x_i^T E(y_i - x_i\beta) = 0 \to E[\nabla l] = \dfrac{1}{\sigma^2}\sum_{i=1}^N x_i^T E(y_i - x_i\beta) = 0$,

   $E[(\nabla l)^2] \to E[(\nabla l)(\nabla l)^T]$

9. P89 Proposition 17: Any covariance matrix $V(\tilde{\beta}) \in \mathbb{R}^{(p+1)\times(p+1)}$ w.r.t. an Unbiased estimate does not exceed the inverse of the Fisher information matrix $\to$ Suppose J is nonsingular. Any covariance matrix $V(\tilde{\beta}) \in \mathbb{R}^{(p+1)\times(p+1)}$ w.r.t. an unbiased estimate is not below the inverse of the Fisher information matrix

10. P91 L3: $\int_S f(x)dx = 0 \Longrightarrow \int_S g(x)dx = 0 \to \int_S f(x)dx > 0 \Longrightarrow \int_S g(x)dx > 0$

11. P91 L8-: the likelihood $\to$ the negated likelihood

12. P92 L11-: Remove ()

13. P93 L13: $\frac{N}{2}\log 2\pi\sigma^2+\frac{1}{2}(p+1)\to\frac{N}{2}\log 2\pi\sigma^2 e+\frac{1}{2}(p+1)$
14. P95 L5: Remove the right $\int_{-\infty}^{\infty}$
15. P96 L1, L3 (one each): $(n-1)$-th $\to$ $(n-j)$-th
16. P96 L6-: $\sigma^2(S)\to\sigma^2$

6 Regularization

1. P100 L6-: the becomes $\to$ the estimate becomes
2. P102 L1: $X^\top X+\lambda I.\to X^\top X+N\lambda I.$
3. P102 In the last lines the table and Figure 6.1 of in Example 48: 18-24 year old in college $\to$ people 25 years + in college
4. P104 L9, L2-: the $\to$ the subderivative
5. P106 L5-: $\displaystyle\frac{1}{n}\sum_{i=1}^n x_{i,j}x_{i,k}=\left\{\begin{array}{ll}1, & j=k\\ 0, & j\neq k\end{array}\right.\to\frac{1}{N}\sum_{i=1}^N x_{i,j}x_{i,k}=\left\{\begin{array}{ll}1, & j=k\\ 0, & j\neq k\end{array}\right.$
6. P105 L2-: $\displaystyle 0\in-\frac{1}{N}\sum_{i=1}^n x_{i,j}\left(y_i-\sum_{k=1}^p x_{i,k}\beta_k\right)\to 0\in-\frac{1}{N}\sum_{i=1}^N x_{i,j}\left(y_i-\sum_{k=1}^p x_{i,k}\beta_k\right)$
7. P107 L1-: $0\in-\frac{1}{N}\sum_{i=1}^n x_{i,j}(r_{i,j}-x_{i,j}\beta_j)\to 0\in-\frac{1}{N}\sum_{i=1}^N x_{i,j}(r_{i,j}-x_{i,j}\beta_j)$
8. P108 L6: Remove `for(j in 1:p){...}`
9. p108 Example 50 and P114 Exercise 114: `seq(0,100,0.1)` $\to$ `seq(0,200,0.1)`, `c(0,100)` $\to$ `c(0,200)`,
10. P110 Figure 6.6: $x\to\beta_1$ and $y\to\beta_2$
11. P112 L9 and P114 Exercise 53 `X=as.matrix(df[,3:7])` $\to$ `X=as.matrix(df[,3:7]), y=df[,1]`

7 Non-Linear Regression

1. P119 L2 Program: `y=round(x)%%2*2-1+rnorm(n)*0.2` $\to$ `y=ceiling(x)%%2*2-1+rnorm(n)*0.2`
2. P120 L8-: for each $i=1,2,\ldots,K+1\to$ for each $i=1,2,\ldots,K$
3. P121 L15-: Remove erty
4. P123 L8-: by specifying $\beta_1,\ldots,\beta,K$ (Fig. 7.6) $\to$ by specifying $\beta_1,\ldots,\beta,K$
5. P124 Example 55: $x_1,x_2,x_3\to\alpha_1,\alpha_2,\alpha_3$
6. P127 L10,L11,L13: $\displaystyle\lambda\int_{-\infty}^{\infty}\{g''(x)\}^2dx\to\lambda\int_{-\infty}^{\infty}f''(x)\}^2dx$, $L(g)\to L(f)$
7. P128 L3+: `for(I in 3:(n))` $\to$ `for(I in 3:n)`

8. P128 Example 57, and Exercise 64: `c(40, 400, 1000)` $\to$ `c(1, 30, 80)`

9. P129 L3 below, and Exercise 65: $H[\lambda] := X^T(X^TX+\lambda G)^{-1}X \to H[\lambda] := X(X^TX+\lambda G)^{-1}X^T$

10. P130 Example 58, and Exercise 65:
    Line 2 `0*1*rnorm(n)` $\to$ `0.1*rnorm(n)`
    Line 4 `function(j,u)` $\to$ `function(j,u,x)`
    Line 5 `function(j,u)` $\to$ `function(j,u,x)`, `d(j-2u)` $\to$ `d(j-2u,x)`, `d(j-2,u)` $\to$ `d(j-2,u,x)`, `d(n-1,u)` $\to$ `d(n-1,u,x)`,

11. P130 L11-: $K \in \mathcal{X}^{n\times n} \to K \in \mathbb{R}^{n\times n}$

12. P131 L8: $\begin{bmatrix} K_\lambda(x_1,y_1) & K_\lambda(x_1,y_2) & K_\lambda(x_1,y_3) \\ K_\lambda(x_2,y_1) & K_\lambda(x_2,y_2) & K_\lambda(x_2,y_3) \\ K_\lambda(x_3,y_1) & K_\lambda(x_3,y_2) & K_\lambda(x_3,y_3) \end{bmatrix} \to \begin{bmatrix} K_\lambda(x_1,x_1) & K_\lambda(x_1,x_2) & K_\lambda(x_1,x_3) \\ K_\lambda(x_2,x_1) & K_\lambda(x_2,x_2) & K_\lambda(x_2,x_3) \\ K_\lambda(x_3,x_1) & K_\lambda(x_3,x_2) & K_\lambda(x_3,x_3) \end{bmatrix}$

13. P131 L13-: $K(x,x_i)$,$K(x,x_j) \to k(x,x_i)$,$k(x,x_j)$

14. P131 L9-: $K(x_*,x_1)$,$K(x_*,x_j)$,$K(x_*,x_N) \to k(x_*,x_1)$,$k(x_*,x_j)$,$k(x_*,x_N)$

15. P133 L12 above: $x \in \mathbb{R} \to x \in \mathbb{R}^p$

16. P134 L14: (7.6) $\to$ (7.7)

17. P134 Example 62: $h_1(x), h_2(x), h_3(x) \to h_3(x), h_4(x), h_5(x)$

18. P135 Example 63, and Exercises 67 and 68: `randn(1)` $\to$ `randn(n)`

19. P136 L3-: $x \le \alpha_k \to x \le \alpha_K$.

20. P137 L8-: fourth $\to$ fofth

21. P139 L6-: first term $\to$ second term

22. P140 L2: second term $\to$ first term

23. Exercise 67: $y_i - \beta(x)^T[1,x_i] \to y_i - [1,x_i]\beta(x)$

8 Decision Trees

1. P148 (8.3): $f(x,y) \to f_{XY}(x,y)$

2. P148 L5-: $x \in R_j \to x_i \in R_j$.

3. P149 L7-: $\{x_k \mid x_k < x_{i,j}\}$, $\{x_k \mid x_k \ge x_{i,j}\} \to \{x_k \mid x_{k,j} < x_{i,j}\}$ and $\{x_k \mid x_{k,j} \ge x_{i,j}\}$.

4. P149 L4-: $\left(y_i - \bar{y}^L_{i,j}\right)^2$, $\left(y_i - \bar{y}^R_{i,j}\right)^2 \to \left(y_k - \bar{y}^L_{k,j}\right)^2$, $\left(y_k - \bar{y}^R_{k,j}\right)^2$

5. P153 L5-: `i=vertex[[h]]$i` $\to$ `th=vertex[[h]]$th`

6. P155 Fig7.5 caption: from 1 to 12 $\to$ from 1 to 15

7. P157 L2: Remove `\bar{y}_j`:

8. P158 Program 7: `T=0; ...z[j])/n` → `T=0; ...z[j])`
9. P160 L8-: `plot(···)` → `plot(···, edge.arrow.size = 0.1)`
10. P160 L14 and Exercise 74: `gbm(···)` → `gbm(···, shrinkage = 0.001)`

9 Support Vector Machine

1. P172 L4: the samples are → the samples are separable
2. P174 L1-: $\alpha' \in \mathbb{R}^m \to \alpha' \geq 0$
3. P175 L17-: $\alpha = -1/\sqrt{2} \to \alpha = 1/\sqrt{2}$
4. P175 L15-(before(9.6)): The problem of minimizing $f(\beta) = \sup_{\alpha \geq 0} L(\alpha, \beta)$ is a primary problem, while that of maximizing $g(\alpha) := \inf_\beta L(\alpha, \beta)$ under $\alpha \geq 0$ is a dual problem. If we write the optimum values of the primary and dual problems as $f^* := \inf_\beta f(\beta)$ and $g^* := \sup_{\alpha \geq 0} g(\alpha)$, then we have
5. P176 L4: $x = x_0 \in \mathbb{R} \to x = x_0 \in \mathbb{R}^p$
6. P176 L7: an arbitrary → each solution
7. P176 L7: $(\beta - \beta^*) \leq f_0(\beta) \to (\beta - \beta^*) = f_0(\beta)$
8. P177 (9.18): $\sum_{i=1}^N \alpha_i y_i x_i \to \sum_{i=1}^N \alpha_i y_i x_i^T$
9. P180 L2: the rather than → the dual rather than the prime
10. P186 Proof of Proposition 25 (three): $\epsilon \to \epsilon_i$
11. P186 L4-, L6-, L9- (one each): $\epsilon \to \epsilon_i$
12. P189 L6-: cannot → can

10 Unsupervised Learning

1. P194 L2-: clustering. In fact → clustering, where C_k is the set of indexes of samples in the k-th cluster. In fact
2. P197 L6 below: L_2 → L2
3. P203 Example 85 Program : insert `n=100` in the first line
4. P205 pca function L4: `t(X)%*%X` → `t(X)%*%X/n`
5. P206 Example 87 (two each): `T[1,2]` → `T[2,1]`, `T[2,1]` → `T[1,2]`
6. P211 Problem 91: `single.complete` → `dist.single`