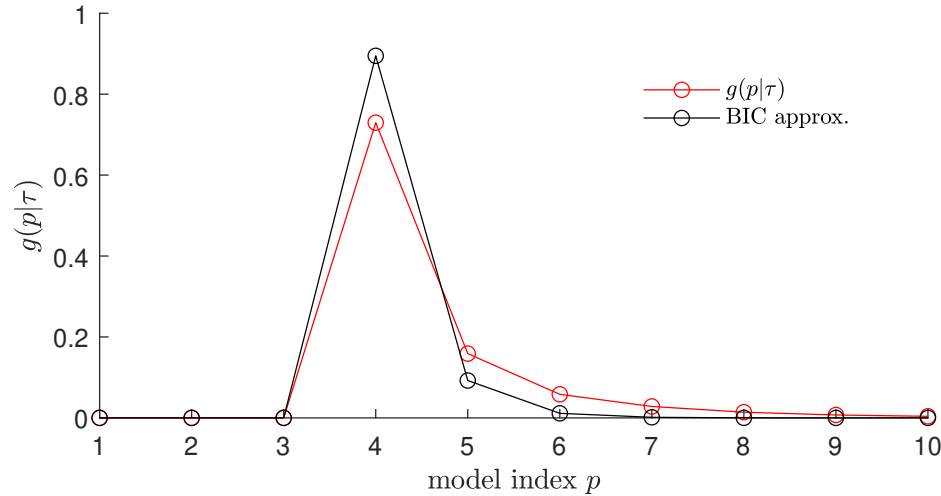


Data Science and Machine Learning: Mathematical and Statistical Methods

Errata

(Last Update 29th April 2022)

1. Page 33, definition of the Hilbert matrix: $\mathbf{H}_p = \int_0^1 [1, u, \dots, u^{p-1}]^\top [1, u, \dots, u^{p-1}] du$.
2. Page 37, line 3 from the top: replace $\mathbb{E}Y_i$ with $\mathbb{E}_{\mathbf{x}}Y_i$.
3. Page 38, lines 3,4 in second paragraph: replace $\ell_{\mathcal{T}_k}$ symbol with ℓ_{C_k} .
4. Page 38, first line in displayed equation: replace $\ell_{\mathcal{T}_k}$ symbol with $\ell_{C_k}(g_{\mathcal{T}_k})$.
5. Page 57, Figure 2.16. There was a mistake in the drawing of the BIC approximation. The actual BIC approximation matches the posterior density quite well:



6. Page 72, Line -2: ... in terms of the probability ... (remove repeated “the”).
7. Page 74, Lines 6 and 10 of `accrejgamma.py`: The parameter `lam` should be replaced with 4 for the proposal pdf `g`.
8. Page 74, Line -3. ... from state x_{t-1} to state x_t ...
9. Page 78, Algorithm 3.27 input: Replace $q(\mathbf{x}, \mathbf{y})$ with $q(\mathbf{y} | \mathbf{x})$.
10. Page 85, Line 7: 0.025 and 0.975 quantiles
11. Page 98, Figure 3.10. Change “unnormalized” to “normalized” in the caption. Also, 0.4,0.2 should be 0.4, 0.2.
12. Page 100, Line -8: $(1 - \alpha v)$ should be $(1 - \alpha)v$.
13. Page 103, line above 4th statement of Algorithm 3.4.4: $X_{(1)}$ should be $\mathbf{X}_{(1)}$.
14. Page 104, Line 2: $\lceil \cdot \rceil$ should be $\lfloor \cdot \rfloor$.

15. Page 108, Lines 5 and 11: $\mathbb{E}S(\lambda)$ should be $S(\lambda)$.
16. Page 109, Line 12: 1.6 should be 0.6.
17. Page 110, Line -2: ... bits that *do not* match ...
18. Page 111, Caption of Figure 3.15: ... that *do not* match ...
19. Page 111, Line 4 under Figure 3.15: “maximize” should be “minimize”
20. Page 112, Lines 1-2: “Note that ...”. Should be deleted.
21. Page 112, Line 3: 200 should be 100.
22. Page 112, Line 8: “maximizer” should be “minimizer”
23. Page 112, Line 12: “maximum” should be “minimum”
24. Page 124, Equation (4.9): $S(X; \theta)$ should be $S(X | \theta)$
25. Page 145, Line 1 of Example 4.6: This refers to Figure 4.4, not Figure 4.8.
26. Page 149, Line -1: $|d_{im} - d_{jm}|$ should be divided by 2.
27. Page 156, Line -5: u_ℓ^\top should be \mathbf{u}_ℓ^\top .
28. Page 151, Line -5: Figure 4.12 depicts the ellipsoid $\mathbf{x}^\top \Sigma^{-1} \mathbf{x} = 1$.
29. Page 160, Exercise 5: $\mathbf{F}(\theta)$ should be $\mathbf{F}(\theta)/n$.
30. Page 162: Line 12: $\Sigma^{1/2} \mathbf{x}$ should be $\Sigma^{-1/2} \mathbf{x}$.
31. Page 162: Lines 17 and 20: $\Sigma^{1/2}(\mathbf{x}_i - \boldsymbol{\mu})$ should be $\Sigma^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu})$.
32. Page 178: fourth line below Table 5.1: replace “qualitative” with “quantitative”.
33. Page 179, Line 5: For independent Y_1, \dots, Y_n , where each Y_i corresponds to the factor values u_{i1}, \dots, u_{ir} , let
34. Page 179, fourth line in Example 5.5: replace “row-wise” with “column-wise” and the vector \mathbf{y} with $\mathbf{y} = [9.2988, 8.2111, 9.0688, 8.2552, 9.4978, \dots, 8.9485]^\top$.
35. Page 179, Line -6. Estimation of $\boldsymbol{\beta}$...
36. Page 181, formula for R_{adjusted}^2 at the bottom: replace $n - p - 1$ in the formula with $n - p$.
37. Page 184, formula for F_i should have the norms squared:
$$F_i = \frac{\|\mathbf{Y}^{(i)} - \mathbf{Y}^{(i-1)}\|^2 / p_i}{\|\mathbf{Y} - \mathbf{Y}^{(d)}\|^2 / (n - p)} .$$
38. Page 211, Exercise 12 (b): \mathbf{P}_{ii} should be $(1 - \mathbf{P}_{ii})$; that is 1 minus the i -th leverage.

39. Page 219, Line -2: ... only β_1 is regularized.
40. Page 221, Line 8: ... one obtains the so-called ...
41. Page 222, 5th line after Definition 6.1: $\kappa(\mathbf{x}, \mathbf{x}')$ should be $\kappa(\mathbf{x}', \mathbf{x})$.
42. Page 235, Line 7: $\int_0^1 (g''(x))^2 dx$ instead of $\int_0^1 (g'')^2 dx$.
43. Page 247, Algorithm 6.8.1, Line 1: \mathbb{R}^p should be \mathbb{R}^n .
44. Page 248, Algorithm 6.8.2, Line 1: Set $\mathbf{B} \leftarrow (n\gamma\mathbf{I}_p)^{-1}$.
45. Page 255, Line 5: Remove “ $\mathbf{L} = \mathbf{I}$ (identity matrix)” , as \mathbf{L} is not the identity matrix.
46. Page 260, equation above (7.13): Put a transpose $^\top$ on the first $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$.
47. Page 261, Line 4 of middle paragraph: In the formula for b , put a transpose $^\top$ on the first $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$.
48. Page 264, Line 8: Replace $g_X(\mathbf{x})$ with $g_X(\mathbf{x}|\boldsymbol{\theta})$
49. Page 273, 3rd line under Figure 7.9: The results are summarized in Table 7.6.
50. Page 288. Replace $x_2 \leq -20.5$ with $x_1 \leq -20.5$ in Figure 8.2 and in the referring text.
51. Page 289, Lines 4–5 of Section 8.2: x_1 should be x_2 .
52. Page 290, first line under Algorithm 8.2.1: change R_{v_T} and R_{v_F} to \mathcal{R}_{v_T} and \mathcal{R}_{v_F} . In the same line, the prediction functions should be g^T and g^F .
53. Page 290, first displayed equation: \mathbb{R}_v should be \mathcal{R}_v .
54. Page 291, line 2: $g^v(\mathbf{x})$ should be $g^w(\mathbf{x})$.
55. Page 308, BaggingExample.py, line starting with “ids = ”: replace `x_test` with `x_train`, twice.
56. Page 310, Algorithm 8.6.1. Delete line 3 and replace line 4 with: Train a decision tree $g_{\mathcal{T}_b^*}$ via Algorithm 8.2.1, where each split is performed using m randomly selected features out of p .
57. Page 313, formula (8.21): g_0 should be $g_0(\mathbf{x})$.
58. Page 316, formula (8.23). Delete $g_{b-1}(\mathbf{x}_i)$.
59. Page 327, warning box: ... j -th node in the $(l-1)$ -st layer with the i -th node in the l -th layer.
60. Page 329, line 2: Section 4.39 should be Equation (4.39).
61. Page 329, line 12 from below: change y_{i-k} to y_{i-k+1} .

62. Page 331, last displayed equation:

$$\frac{\partial C}{\partial \mathbf{b}_l} = \frac{\partial z_l}{\partial \mathbf{b}_l} \frac{\partial C}{\partial z_l} = \delta_l, \quad l = 1, \dots, L.$$

63. Page 333, line 4 of Example 9.4: “inputs y ” should be “inputs \mathbf{x} ”.

64. Page 333, lines –8, –7: θ should be $\boldsymbol{\theta}$ (three times).

65. Page 335, Algorithm 9.4.2, Line 2: ... using $\frac{\partial C}{\partial \mathbf{g}} = 1$...

66. Page 339, Equation (9.10): the summation should go to t , not h . On the next line \mathbf{u}_k should also be changed to \mathbf{u}_t .

67. Page 340, second displayed line:

$$[p_0, p_1, p_2, p_3] = [1, 20, 20, 1].$$

68. Page 341, Line 3: Remove the line $S = \text{RELU}$.

69. Page 351, Exercise 7(b): In the displayed formula, \mathbf{B} should be replaced with \mathbf{B}^{-1} .

70. Page 361, 3rd line in the proof of Theorem A.3: $\{v_i\}$ should be $\{\mathbf{v}_i\}$.

71. Page 362, First sentence in paragraph above Theorem A.4: ... the matrix \mathbf{P} projects any vector in \mathcal{V} onto itself.

72. Page 362, Sentence above Theorem A.4: ... where \mathbf{U} is not ...

73. Page 376, last displayed equation: Replace $[x_1, x_2]$ with $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

74. Page 377, Lines 6,7: \mathbf{u}_n should be \mathbf{u}_m and \mathbb{C}^n should be \mathbb{C}^m .

75. Page 378, line 3: Replace \mathbf{V}^\top with \mathbf{V} .

76. Page 380, third line from below: change b_{i-k} to b_{i-k+1} .

77. Page 386, equation A.35: replace $\sum_{j=0}^{\infty} \frac{\sin(2j+1)}{2j+1}$ with $\sum_{k=1}^{\infty} \frac{\sin(kx)}{k}$.

78. Page 394, line 5: ... can be computed with the aid ... (missing “the”)

79. Page 385, halfway: ... counting measure on \mathbb{Z}^d .

80. Page 400, first two lines of Section B.1.2: $\mathbf{f} : \mathbb{R}^k \rightarrow \mathbb{R}^m$, $\mathbf{g} : \mathbb{R}^m \rightarrow \mathbb{R}^n$, and $\mathbf{g} \circ \mathbf{f}$ is a function from \mathbb{R}^k to \mathbb{R}^n .

81. Page 401, Line –1: By Theorem A.8 ...

82. Page 405, Line 8: replace “ $f(\mathbf{x} + t_1 \mathbf{d}) \geq f(\mathbf{x}) + t_1 \mathbf{v}^\top \mathbf{d}$ and $f(\mathbf{x} + t_2 \mathbf{d}) \geq f(\mathbf{x}) + t_2 \mathbf{v}^\top \mathbf{d}$ for some subgradient \mathbf{v} ” with “ $f(\mathbf{a}) \geq f(\mathbf{b}) + (\mathbf{a} - \mathbf{b})^\top \mathbf{v}$ for some subgradient \mathbf{v} ” and replace “Subtracting the last two equations” with “Substituting with $\mathbf{a} = \mathbf{x} + t_1 \mathbf{d}$ and $\mathbf{b} = \mathbf{x} + t_2 \mathbf{d}$ ”.

83. Page 404, last two lines: replace H with \mathbf{H} .
84. Page 414, Section B.3.4: Replace ℓ with ℓ_τ .
85. Page 417, line 2 of 3-rd paragraph: replace i with j in “the equalities $g_j(\mathbf{x}) + s_j = 0$ for all i .”
86. Page 418, line 2: replace \mathbf{B}^\top with \mathbf{B} .
87. Page 433, displayed equation in the proof of Theorem C.4: replace $|\mathbf{J}_{\mathbf{g}^{-1}}(\mathbf{z})|$ with $|\det(\mathbf{J}_{\mathbf{g}^{-1}}(\mathbf{z}))|$.
88. Page 439, line 4: is equal to $\Gamma(\alpha)\lambda^{-\alpha}$ times ...
89. Page 442, 4th line from the bottom: $x \geq c$ should be $x > c$.
90. Page 445, halfway on the page: $|e^{ix} - 1| = \left| \int_0^x i e^{i\theta} d\theta \right| \leq \left| \int_0^x |i e^{i\theta}| d\theta \right| = |x|$.
91. Page 446, displayed equation below (C.37): $O(t/n)$ should be $o(t/n)$, and in the next displayed equation, $o(1)$ should be $o(1/n)$.
92. Page 448, line 2: $O(t^3/n^{3/2})$ should be $o(t^2/n)$.
93. Page 450, first displayed equation after (C.39): The Σ in the denominator should be Σ_n .
94. Page 451: Delete “ln” after “An application ... yields”
95. Page 451, line starting with “asymptotically negligible”: Replace n with $-n$ in the exponent.
96. Page 456, Sentence under (C.47): Similar to the one-dimensional case ($d = 1$), replacing the factor $1/n$ with $1/(n-1)$ gives an unbiased estimator, called the *sample covariance matrix*.
97. Page 457, last line of Example C.13: $g'(\theta)$ should be $l'(\theta)$.
98. Page 511, line 13 from above: ‘expectation of’.