Data Science and Machine Learning:
Mathematical and Statistical Methods

Errata

(Last Update 30th October 2023)

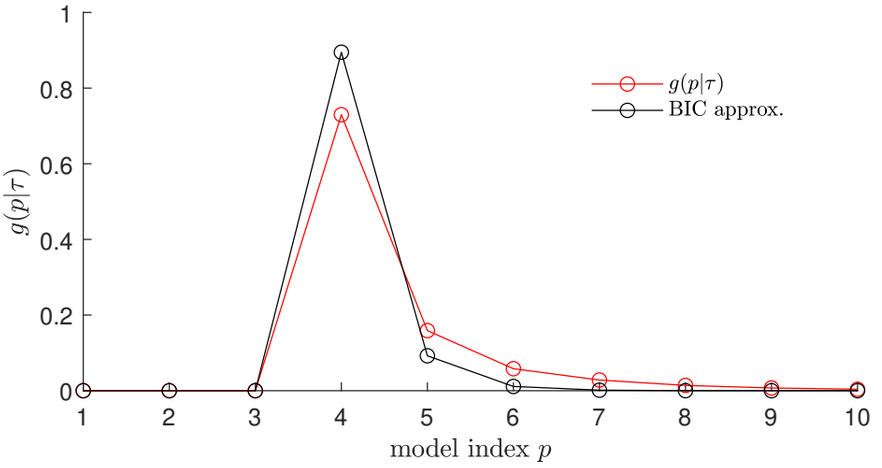
1. Page 33, definition of the Hilbert matrix: $\mathbf{H}_p = \int_0^1 [1, u, \ldots, u^{p-1}]^\top [1, u, \ldots, u^{p-1}]\, \mathrm{d}u$.

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 $q(\boldsymbol{x}, \boldsymbol{y})$ with $q(\boldsymbol{y} \,|\, \boldsymbol{x})$.

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: $(1 - \alpha \boldsymbol{v})$ should be $(1 - \alpha)\boldsymbol{v}$.

13. Page 103, line above 4th statement of Algorithm 3.4.4: $X_{(1)}$ should be $\mathbf{X}_{(1)}$.

14. Page 104, Line 2: $\lceil\ \rceil$ should be $\lfloor\ \rfloor$.

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): $\boldsymbol{S}(\boldsymbol{X};\boldsymbol{\theta})$ should be $\boldsymbol{S}(\boldsymbol{X}\,|\,\boldsymbol{\theta})$

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: $u_\ell^\top$ should be $\boldsymbol{u}_\ell^\top$.

28. Page 151, Line −5: Figure 4.12 depicts the ellipsoid $\boldsymbol{x}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{x} = 1$.

29. Page 160, Exercise 5: $\mathbf{F}(\boldsymbol{\theta})$ should be $\mathbf{F}(\boldsymbol{\theta})/n$.

30. Page 162: Line 12: $\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}$ should be $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}$.

31. Page 162: Lines 17 and 20: $\boldsymbol{\Sigma}^{1/2}(\boldsymbol{x}_i - \boldsymbol{\mu})$ should be $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}_i - \boldsymbol{\mu})$.

32. Page 178: fourth line below Table 5.1: replace "qualitative" with "quantitative".

33. Page 179, Line 5: For independent $Y_1, \ldots, Y_n$, where each Y_i corresponds to the factor values $u_{i1}, \ldots, u_{ir}$, let

34. Page 179, fourth line in Example 5.5: replace "row-wise" with "column-wise" and the vector $\boldsymbol{y}$ with $\boldsymbol{y} = [9.2988, 8.2111, 9.0688, 8.2552, 9.4978, \ldots, 8.9485]^\top$.

35. Page 179, Line −6. Estimation of $\boldsymbol{\beta}$ . . .

36. Page 181, formula for R^2_{adjusted} 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:

$$F_i = \frac{\|\boldsymbol{Y}^{(i)} - \boldsymbol{Y}^{(i-1)}\|^2/p_i}{\|\boldsymbol{Y} - \boldsymbol{Y}^{(d)}\|^2/(n-p)}\,.$$

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: $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ should be $\kappa(\boldsymbol{x}', \boldsymbol{x})$.

42. Page 235, Line 7: $\int_0^1 (g''(x))^2 \, \mathrm{d}x$ instead of $\int_0^1 (g'')^2 \, \mathrm{d}x$.

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 $g_X(\boldsymbol{x})$ with $g_X(\boldsymbol{x} \,|\, \boldsymbol{\theta})$

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 $\mathcal{R}_{v_\mathrm{T}}$ and $\mathcal{R}_{v_\mathrm{F}}$. 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: $g^v(\boldsymbol{x})$ should be $g^w(\boldsymbol{x})$.

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 $g_0(\boldsymbol{x})$.

58. Page 316, formula (8.23). Delete $g_{b-1}(\boldsymbol{x}_i)$.

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:

$$\frac{\partial C}{\partial \boldsymbol{b}_l} = \frac{\partial \boldsymbol{z}_l}{\partial \boldsymbol{b}_l}\frac{\partial C}{\partial \boldsymbol{z}_l} = \boldsymbol{\delta}_l, \quad l = 1, \ldots, L.$$

63. Page 333, line 4 of Example 9.4: "inputs y" should be "inputs $\boldsymbol{x}$".

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

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

66. Page 336, the secant condition $\mathbf{C}_t \boldsymbol{g}_{t-1} = \boldsymbol{\delta}_{t-1}$ should read $\mathbf{C}_t \boldsymbol{g}_t = \boldsymbol{\delta}_t$, and the definitions on the next line should read $\boldsymbol{\delta}_t := \boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}$ and $\boldsymbol{g}_t := \boldsymbol{u}_t - \boldsymbol{u}_{t-1}$.

67. Page 336, Equation (9.7): replace $\mathbf{C}_{t+1}$ with $\mathbf{C}_t$ and $\mathbf{C}_t$ with $\mathbf{C}_{t-1}$ to obtain:

$$\mathbf{C}_t = \left(\mathbf{I} - \upsilon_t \boldsymbol{g}_t \boldsymbol{\delta}_t^\top\right)^\top \mathbf{C}_{t-1} \left(\mathbf{I} - \upsilon_t \boldsymbol{g}_t \boldsymbol{\delta}_t^\top\right) + \upsilon_t \boldsymbol{\delta}_t \boldsymbol{\delta}_t^\top, \quad \upsilon_t := (\boldsymbol{g}_t^\top \boldsymbol{\delta}_t)^{-1}.$$

68. Page 337, Equation (9.9): replace i with j two times.

69. Page 337, line 3: change "At each iteration" to "At the final iteration".

70. Page 337, line 5: change "By iterating this recursion" to "By iterating the recursion (9.7)".

71. Page 338, Algorithm 9.4.3, second line from top: the equation $\mathbf{C}_h = \left(\mathbf{I} - \upsilon_j \boldsymbol{\delta}_j \boldsymbol{g}_j^\top\right) \mathbf{C}_{t-1} \left(\mathbf{I} - \upsilon_j \boldsymbol{g}_j \boldsymbol{\delta}_j^\top\right) + \upsilon_t \boldsymbol{\delta}_t \boldsymbol{\delta}_t^\top$ should read $\mathbf{C}_t = (\mathbf{I} - \upsilon_t \boldsymbol{\delta}_t \boldsymbol{g}_t^\top) \mathbf{C}_{t-1} \left(\mathbf{I} - \upsilon_t \boldsymbol{g}_t \boldsymbol{\delta}_t^\top\right) + \upsilon_t \boldsymbol{\delta}_t \boldsymbol{\delta}_t^\top$.

72. Page 338, line 2 from below: identities should read $\boldsymbol{\delta}_t = \boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}$ and $\boldsymbol{g}_t = \boldsymbol{u}_t - \boldsymbol{u}_{t-1}$.

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

74. Page 339, Algorithm 9.4.5: in lines 3 and 4 replace the first appearance of $\hat{\boldsymbol{u}}_t$ and $\boldsymbol{v}_t$ with $\boldsymbol{u}_t^*$ and $\boldsymbol{v}_t^*$, respectively; then, in line 5, replace $\hat{\boldsymbol{u}}_t$ and $\boldsymbol{v}_t$ with $\boldsymbol{u}_t^*$ and $\boldsymbol{v}_t^*$, respectively.

75. Page 340, line 3 and 4 from above: Replace $\boldsymbol{\delta}_{t-1}$ with $\boldsymbol{\delta}_t$ two times.

76. Page 340, second displayed line:

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

77. Page 341, Line 3: Remove the line `S = RELU`.

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

79. Page 353, in Exercise 8: replace $\mathbf{C}_{t+1}$ with $\mathbf{C}_t$ and $\mathbf{C}_t$ with $\mathbf{C}_{t-1}$.

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

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

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

83. Page 374, Line 17: $l_{kk} = \sqrt{a_{kk} - \|\boldsymbol{l}_k\|^2}$ should be $l_{kk} = \sqrt{a_{kk} - \|\boldsymbol{l}_{k-1}\|^2}$.

84. Page 376, last displayed equation: Replace $\left[\boldsymbol{x}_1, x_2\right]$ with $\begin{bmatrix}\boldsymbol{x}_1 \\ x_2\end{bmatrix}$.

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

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

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

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

89. Page 394, line 5: ... can be computed with the aid ... (missing "the")

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

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

92. Page 401, Line −1: By Theorem A.8 ...

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

94. Page 404, last two lines: replace H with $\mathbf{H}$.

95. Page 406, Table B.1, Linear program constraints, replace $\boldsymbol{x} \geq 0$ with $\boldsymbol{x} \geq \mathbf{0}$.

96. Page 414, Section B.3.4: Replace ℓ with ℓ_τ.

97. Page 417, line 2 of 3-rd paragraph: replace i with j in "the equalities $g_j(\boldsymbol{x}) + s_j = 0$ for all i."

98. Page 418, line 2: replace $\mathbf{B}^\top$ with $\mathbf{B}$.

99. Page 433, displayed equation in the proof of Theorem C.4: replace $|\mathbf{J}_{\mathbf{g}^{-1}}(z)|$ with $|\det(\mathbf{J}_{\mathbf{g}^{-1}}(z))|$.

100. Page 439, line 4: is equal to $\Gamma(\alpha)\lambda^{-\alpha}$ times ...

101. Page 442, 4th line from the bottom: $x \geq c$ should be $x > c$.

102. Page 445, halfway on the page: $|\mathrm{e}^{\mathrm{i}x} - 1| = \left|\int_0^x \mathrm{i}\,\mathrm{e}^{\mathrm{i}\theta}\,\mathrm{d}\theta\right| \leq \left|\int_0^x |\mathrm{i}\,\mathrm{e}^{\mathrm{i}\theta}|\,\mathrm{d}\theta\right| = |x|$.

103. 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 $o(1/n)$.

104. Page 448, line 2: $\dot{O}(t^3/n^{3/2})$ should be $o(t^2/n)$.

105. Page 450, first displayed equation after (C.39): The $\mathbf{\Sigma}$ in the denominator should be $\mathbf{\Sigma}_n$.

106. Page 451: Delete "ln" after "An application ... yields"

107. Page 451, line starting with "asymptotically negligible": Replace n with $-n$ in the exponent.

108. 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*.

109. Page 457, last line of Example C.13: $g'(\theta)$ should be $l'(\theta)$.

110. Page 511, line 13 from above: 'expectation of'.