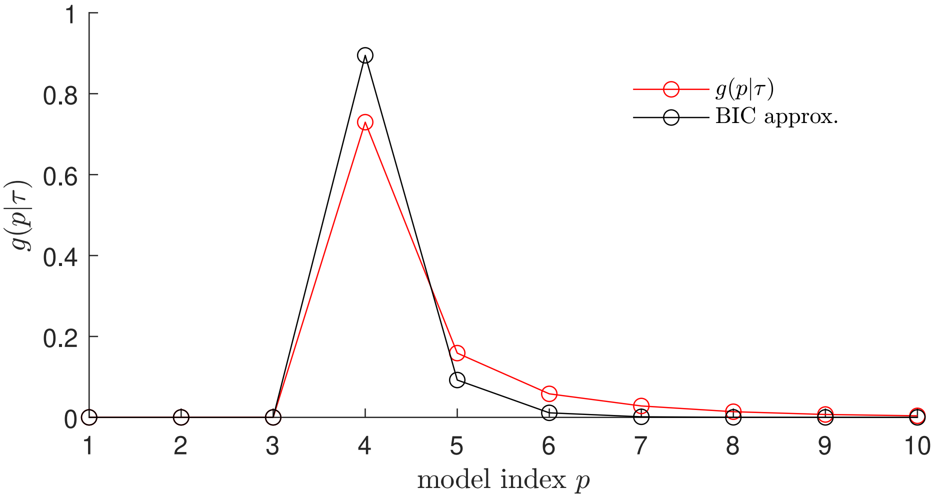
Data Science and Machine Learning:
Mathematical and Statistical Methods

Errata

(Last Update 25th August 2021)

1. Page 33, definition of the Hilbert matrix: $\mathbf{H}_p = \int_0^1 [1, u, \dots, u^{p-1}]^\top [1, u, \dots, u^{p-1}] \, \mathrm{d}u$.
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 100, Line −8: $(1 - \alpha\boldsymbol{v})$ should be $(1 - \alpha)\boldsymbol{v}$.
11. Page 108, Lines 5 and 11: $\mathbb{E}S(\lambda)$ should be $S(\lambda)$.
12. Page 162: Line 12: $\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}$ should be $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}$.
13. Page 162: Lines 17 and 20: $\boldsymbol{\Sigma}^{1/2}(\boldsymbol{x}_i - \boldsymbol{\mu})$ should be $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}_i - \boldsymbol{\mu})$.
14. Page 178: fourth line below Table 5.1: replace "qualitative" with "quantitative".

15. Page 179, fourth line in Example 5.5: replace "row-wise" with "column-wise" and the vector $\boldsymbol{y}$ with $\boldsymbol{y} = [9.2988, 8.2111, 9.0688, 8.2552, 9.4978, \ldots, 8.9485]^\top$.
16. Page 181, formula for R^2_{adjusted} at the bottom: replace $n - p - 1$ in the formula with $n - p$.
17. Page 184, formula for F_i should have the norms squared:

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

18. Page 211, Exercise 12 (b): $\mathbf{P}_{ii}$ should be $(1 - \mathbf{P}_{ii})$; that is 1 minus the i-th leverage.
19. Page 219, Line -2: …only β_1 is regularized.
20. Page 221, Line 8: … one obtains the so-called …
21. Page 235, Line 7: $\int_0^1 (g''(x))^2 \, \mathrm{d}x$ instead of $\int_0^1 (g'')^2 \, \mathrm{d}x$.
22. Page 247, Algorithm 6.8.1, Line 1: $\mathbb{R}^p$ should be $\mathbb{R}^n$.
23. Page 248, Algorithm 6.8.2, Line 1: Set $\mathbf{B} \leftarrow (n\gamma \mathbf{I}_p)^{-1}$.
24. Page 264, Line 8: Replace $g_{\boldsymbol{X}}(\boldsymbol{x})$ with $g_{\boldsymbol{X}}(\boldsymbol{x} \,|\, \boldsymbol{\theta})$
25. Page 273, 3rd line under Figure 7.9: The results are summarized in Table 7.6.
26. Page 290, first line under Algorithm 8.2.1: change R_{ν_T} and R_{ν_F} to $\mathcal{R}_{\nu_\mathrm{T}}$ and $\mathcal{R}_{\nu_\mathrm{F}}$.
27. Page 291, line 2: $g^v(\boldsymbol{x})$ should be $g^w(\boldsymbol{x})$.
28. Page 313, formula (8.21): g_0 should be $g_0(\boldsymbol{x})$.
29. Page 329, line 12 from below: change y_{i-k} to y_{i-k+1}.
30. Page 331, last displayed equation:

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

31. Page 333, line 4 of Example 9.4: "inputs y" should be "inputs $\boldsymbol{x}$".
32. Page 335, Algorithm 9.4.2, Line 2: …using $\frac{\partial C}{\partial \boldsymbol{g}} = 1$ …
33. Page 340, second displayed line:

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

34. Page 341, Line 3: Remove the line `S = RELU`.
35. Page 351, Exercise 7(b): In the displayed formula, $\mathbf{B}$ should be replaced with $\mathbf{B}^{-1}$.

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

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

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

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

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

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

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

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

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

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

46. Page 446, displayed equation below (C.37): $\mathcal{O}(t/n)$ should be $o(t/n)$, and in the next displayed equation, $o(1)$ should be $o(1/n)$.

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

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

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

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

51. 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*.

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

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