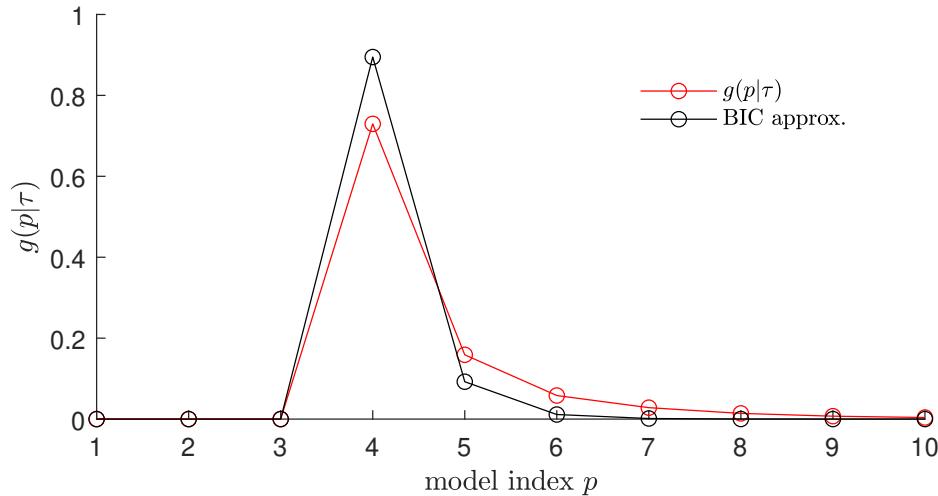


# Data Science and Machine Learning: Mathematical and Statistical Methods

## Errata

(Last Update 26th August 2021)

1. Page 33, definition of the Hilbert matrix:  $\mathbf{H}_p = \int_0^1 [1, u, \dots, u^{p-1}]^\top [1, u, \dots, u^{p-1}] du.$
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  $\ell_{C_k}$ .
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 100, Line -8:  $(1 - \alpha v)$  should be  $(1 - \alpha)v$ .
11. Page 108, Lines 5 and 11:  $\mathbb{E}S(\lambda)$  should be  $S(\lambda)$ .
12. Page 110, Line -2: ... bits that *do not* match ...
13. Page 111, Caption of Figure 3.15: ... that *do not* match ...
14. Page 111, Line 4 under Figure 3.15: “maximize” should be “minimize”
15. Page 112, Lines 1-2: “Note that ...”. Should be deleted.

16. Page 112, Line 3: 200 should be 100.
17. Page 112, Line 8: “maximizer” should be “minimizer”
18. Page 112, Line 12: “maximum” should be “minimum”
19. Page 162: Line 12:  $\Sigma^{1/2}\mathbf{x}$  should be  $\Sigma^{-1/2}\mathbf{x}$ .
20. Page 162: Lines 17 and 20:  $\Sigma^{1/2}(\mathbf{x}_i - \boldsymbol{\mu})$  should be  $\Sigma^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu})$ .
21. Page 178: fourth line below Table 5.1: replace “qualitative” with “quantitative”.
22. Page 179, fourth line in Example 5.5: replace “row-wise” with “column-wise” and the vector  $\mathbf{y}$  with  $\mathbf{y} = [9.2988, 8.2111, 9.0688, 8.2552, 9.4978, \dots, 8.9485]^\top$ .
23. Page 181, formula for  $R_{\text{adjusted}}^2$  at the bottom: replace  $n - p - 1$  in the formula with  $n - p$ .
24. Page 184, formula for  $F_i$  should have the norms squared:
$$F_i = \frac{\|\mathbf{Y}^{(i)} - \mathbf{Y}^{(i-1)}\|^2/p_i}{\|\mathbf{Y} - \mathbf{Y}^{(d)}\|^2/(n-p)}.$$
25. Page 211, Exercise 12 (b):  $\mathbf{P}_{ii}$  should be  $(1 - \mathbf{P}_{ii})$ ; that is 1 minus the  $i$ -th leverage.
26. Page 219, Line –2: ... only  $\beta_1$  is regularized.
27. Page 221, Line 8: ... one obtains the so-called ...
28. Page 235, Line 7:  $\int_0^1 (g''(x))^2 dx$  instead of  $\int_0^1 (g'')^2 dx$ .
29. Page 247, Algorithm 6.8.1, Line 1:  $\mathbb{R}^p$  should be  $\mathbb{R}^n$ .
30. Page 248, Algorithm 6.8.2, Line 1: Set  $\mathbf{B} \leftarrow (n\gamma\mathbf{I}_p)^{-1}$ .
31. Page 264, Line 8: Replace  $g_X(\mathbf{x})$  with  $g_X(\mathbf{x} | \boldsymbol{\theta})$
32. Page 273, 3rd line under Figure 7.9: The results are summarized in Table 7.6.
33. Page 290, first line under Algorithm 8.2.1: change  $R_{v_T}$  and  $R_{v_F}$  to  $\mathcal{R}_{v_T}$  and  $\mathcal{R}_{v_F}$ .
34. Page 291, line 2:  $g^v(\mathbf{x})$  should be  $g^w(\mathbf{x})$ .
35. Page 313, formula (8.21):  $g_0$  should be  $g_0(\mathbf{x})$ .
36. Page 329, line 12 from below: change  $y_{i-k}$  to  $y_{i-k+1}$ .
37. Page 331, last displayed equation:
$$\frac{\partial C}{\partial \mathbf{b}_l} = \frac{\partial \mathbf{z}_l}{\partial \mathbf{b}_l} \frac{\partial C}{\partial \mathbf{z}_l} = \boldsymbol{\delta}_l, \quad l = 1, \dots, L.$$
38. Page 333, line 4 of Example 9.4: “inputs  $y$ ” should be “inputs  $\mathbf{x}$ ”.

39. Page 335, Algorithm 9.4.2, Line 2: ... using  $\frac{\partial C}{\partial g} = 1 \dots$
40. Page 340, second displayed line:
- $$[p_0, p_1, p_2, p_3] = [1, 20, 20, 1].$$
41. Page 341, Line 3: Remove the line  $S = \text{RELU}$ .
42. Page 351, Exercise 7(b): In the displayed formula,  $\mathbf{B}$  should be replaced with  $\mathbf{B}^{-1}$ .
43. Page 362, First sentence in paragraph above Theorem A.4: ... the matrix  $\mathbf{P}$  projects any vector in  $\mathcal{V}$  onto itself.
44. Page 362, Sentence above Theorem A.4: ... where  $\mathbf{U}$  is not ...
45. Page 380, third line from below: change  $b_{i-k}$  to  $b_{i-k+1}$ .
46. Page 394, line 5: ... can be computed with the aid ... (missing “the”)
47. Page 404, last two lines: replace  $H$  with  $\mathbf{H}$ .
48. Page 414, Section B.3.4: Replace  $\ell$  with  $\ell_\tau$ .
49. Page 433, displayed equation in the proof of Theorem C.4: replace  $|\mathbf{J}_{g^{-1}}(\mathbf{z})|$  with  $|\det(\mathbf{J}_{g^{-1}}(\mathbf{z}))|$ .
50. Page 439, line 4: is equal to  $\Gamma(\alpha)\lambda^{-\alpha}$  times ...
51. Page 442, 4th line from the bottom:  $x \geq c$  should be  $x > c$ .
52. Page 445, halfway on the page:  $|e^{ix} - 1| = \left| \int_0^x i e^{i\theta} d\theta \right| \leq \left| \int_0^x |i e^{i\theta}| d\theta \right| = |x|$ .
53. 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  $o(1/n)$ .
54. Page 448, line 2:  $O(t^3/n^{3/2})$  should be  $o(t^2/n)$ .
55. Page 450, first displayed equation after (C.39): The  $\Sigma$  in the denominator should be  $\underline{\Sigma}_n$ .
56. Page 451: Delete “In” after “An application ... yields”
57. Page 451, line starting with “asymptotically negligible”: Replace  $n$  with  $-n$  in the exponent.
58. 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*.
59. Page 457, last line of Example C.13:  $g'(\theta)$  should be  $l'(\theta)$ .
60. Page 511, line 13 from above: ‘expectation of’.