

# Murphy-Machine-Learning-mathematical-errata

Kai Song\*

(Dated: Jan, 2018)

---

\* ks838@cam.ac.uk

@Author:

Kai Song, PhD candidate in quantum dynamics

ks838 at cam.ac.uk

The new errata can be downloaded from my Github: <https://github.com/ks838>

@Notes:

1. In this project, I listed the printing and mathematical typos I found in the 4th printing (the 5th, 6th are the same as the 4th printing). I wish this would be helpful for other readers.
2. A drawback, in my view, is that some of the notations and notes of the figures in the book tend to be explained not very clearly. The axes are often not labeled.

@Introduction to the book:

1. This is a great book, which I have spent a lot of effort on it. I think it has a similar style as the most famous Course of Theoretical Physics by Landau and Lifshitz.
2. There is only one edition till now (Jan, 2018). As the author Kevin Murphy said (I emailed him), the 2nd edition would appear at the end of 2018. We believe the second edition would include more on deep learning and reinforcement learning.
3. website of this book: <http://www.cs.ubc.ca/~murphyk/MLbook/index.html>

## I. CHAP 1

Not found.

## II. CHAP 2

**2-1.** Page 32, Section 2.2.5:

I don't think the introduction of  $f(x)$  is necessary. I mean,  $f(x)$  in Eq. (2.20) and  $p(x)$  in Eq.(2.21) are kind of confusing.

**2-2.** Page 45, Section 2.5.1

'Covariances can be between 0 and infinity' should be 'Covariances can be between  $(-\infty, \infty)$ '.

**2-3.** Page 48, Section 2.5.1

Figure 2.14(c) ' $\alpha = (20, 2, 2)$ ' should be ' $\alpha = (20, 20, 20)$ '.

### III. CHAP 3

**3-1.** Page 74, Section 3.2.4, 2nd paragraph

'so only 4 and 16 get ...' should be 'so only 4, 16 and 64 get ...'

**3-2.** Page 81, Section 3.4, Eq. (3.37)

$x \in S_K$  should be  $\theta \in S_K$

**3-3.** Page 84, Section 3.4.4.1, Eq. (3.52)

17 should be 16

**3-4.** Page 84, Section 3.4.4.1, Eq. (3.53)

27 should be 26, and the last one 5/27 should be 4/26.

**3-5.** Page 87, Section 3.5.2, Eq. (3.65)

$d\theta_{jc}$  should be added.

### IV. CHAP 4

**4-1** Page 106, Section 4.2.2, Eq. (4.43)

$x^T(\beta_{c'} - \beta_c) = \gamma_{c'} - \gamma_c$  should be  $x^T(\beta_c - \beta_{c'}) = \gamma_{c'} - \gamma_c$

**4-2** Page 108, Section 4.2.4, Eq. (4.52)

The rightmost,  $\mathcal{N}(x|\mu_c, \Sigma_c)$  should be  $\mathcal{N}(x_i|\mu_c, \Sigma_c)$

**4-3** Page 113, in Eq.(4.70)

$\rho$  is the correlation coefficient

$$\rho \triangleq \frac{\sigma_{12}}{\sigma_1 \sigma_2} \quad (1)$$

**4-4** at the end of Page 135, below Eq.(4.208)

$\mu_0 = \bar{x}$  should be  $m_0 = \bar{x}$

### V. CHAP 5

**5-1** at the end of Page 154, Section 5.2.2, below Eq. (5.8)

$(1 - \alpha)/2$  should be  $\alpha/2$

**5-2** Page 168, Section 5.4.2, in Eq. (5.52)

$\left(\frac{d \log p(X|\phi)}{d\phi}\right) 2$  should be  $\left(\frac{d \log p(X|\phi)}{d\phi}\right)^2$

**5-3** Page 172, Section 5.4.4.1, Eq. (5.73)

The rightmost  $\text{Beta}(\theta|50, 20)$  should be  $\text{Beta}(\theta|30, 20)$

**5-4** Page 182, Section 5.7.1.5, Eq.(5.109)

in the middle,  $[l(y, \delta(x))]$  should be  $[L(y, \delta(x))]$  for consistence.

## VI. CHAP 6

**6-1** Page 205, Section 6.4.4.1, Eq.(6.45)

$\omega\theta_0$  should be  $\omega\theta^*$

**6-2** Page 218, Exercise 6.3

A ] is left.

## VII. CHAP 7

**7-1** Page 221, Section 7.3.1, Eq. (7.11)

The original one is chaotic, I just rewrote as

$$\frac{1}{2}\text{RSS}(\mathbf{w}) = \frac{1}{2}(\mathbf{y} - \mathbf{X}\mathbf{w})^T(\mathbf{y} - \mathbf{X}\mathbf{w}) = \frac{1}{2}\mathbf{w}^T(\mathbf{X}^T\mathbf{X})\mathbf{w} - \mathbf{w}^T(\mathbf{X}^T\mathbf{y}) + \mathbf{w} - \text{free terms}$$

**7-2** Page 228, Section 7.5.1, Eq. (7.32)

$\frac{1}{N}$  should be deleted.

**7-3** Page 234, Section 7.6.1, below Eq. (7.59)

3rd line under Eq. (7.59): the superscript  $(s)$  is unnecessary.  $\mathbf{w}^{(s)}$  should be  $\mathbf{w}$ .

## VIII. CHAP 8

**8-1** Page 258, Section 8.4.2, in Eq. (8.56)

$N^d$  should be  $N^D$ .

**8-2** Page 265, Section 8.5.1, under Eq. (8.78)

$\text{proj}_{\mathcal{V}}(\mathbf{V}) = \text{argmin}_{\mathbf{W} \in \mathcal{V}} \|\mathbf{W} - \mathbf{V}\|_2$ : the notations are kind of perplexing.

**8-3** Page 277, Section 8.6.3.2, in Eq. (8.115)

$(z_i - m)(z_i - m)^T$  should be  $(m_c - m)(m_c - m)^T$

**8-4** Page 279, Exercise 8.1

Line starts with 'Feature 56', 'max is 45.0, mean is 52.6' should be 'max is 52.6, mean is 45.0 '.

## IX. CHAP 9

**9-1** Page 286, Section 9.2.2.3 Eq.(9.22)

$\exp(\theta^T \phi(x))$  should be  $\exp(\theta^T \phi(x) - A(\theta))$

**9-2** Page 293, Section 9.3.1, below Eq.(9.80)

$\mu = \text{sigm}(\eta)$  should be  $\eta = \text{sigm}(\mu)$

## X. CHAP 10

**10-1** Page 320, Section 10.2.5, under Eq.(10.18)  $t > s$  should  $t < s$ .

## XI. CHAP 11

**11-1** Page 351, Section 11.4, Eq.(11.16)

It should be

$$\text{NLL}(\theta) \triangleq -\frac{1}{N} \log p(\mathcal{D}|\theta)$$

**11-2** Page 360, Section 11.4.4, Eq. (11.56)

$x_{it} = i$  should be  $x_{it} = k$ .

**11-3** Page 372, Section 11.5.1, first paragraph, It should be  $\text{argmax}_{\mathbf{K}} p(\mathcal{D}|\mathbf{K})$ .

## XII. CHAP 12

**12-1** Page 384, Figure 12.1 is really annoying, which may need further explanation.

**12-2** Page 393, Section 12.2.2, below Eq. (12.37)

It should be  $\frac{1}{N} \sum_{i=1}^N x_i x_i^T$

**12-3** Page 397, Section 12.2.4, Theorem 12.2.2,

'S' should be  $\Sigma$ .

**12-4** Page 397, Section 12.2.4, Theorem 12.2.2, Eq. (12.60)

It should be

$$\log p(X|W, \sigma^2) = -\frac{N}{2} \left( \ln |C| - \frac{1}{2} \sum_{i=1}^N x_i^T C^{-1} x_i \right) = -\frac{N}{2} (\ln |C| + \text{tr}(C^{-1} \Sigma))$$

**12-5** Page 404, Section 12.4, around Eq.(12.74)

There is an abused use of the subscript  $j$  and  $r$ .

### XIII. CHAP 13

**13-1** Page 434, Section 13.3.2, Eq.(13.43)

It should be  $f(W) = \text{RSS}(W) + \lambda \|W\|_1$ .

### XIV. CHAP 14

**14-1** Page 483, Section 14.2.3, around Eq.(14.11)

$U_{:j}$  should be  $U_{:,j}$ .

**14-2** Page 484, Section 14.2.5

I think further explanation is needed for the parameters in Matern kernel.

And ,the discussion of Ornstein-Uhlenbeck process on the next page may be confusing.

The original one sounds like Wiener process.

**14-3** Page 499, Section 14.5, in Figure 14.10

An abused use of  $\xi$ . They should changed to  $\xi^+$  and  $\xi^-$  according to Eq. (14.50).

**14-4**

### XV. CHAP 15

**15-1** Page 537, Section 15.4.5, above Eq. (15.92)

$\mathcal{N}(\nu_j|0, \sigma_\omega^2)$  should be  $\mathcal{N}(\nu_j|0, \sigma_\nu^2)$

### XVI. CHAP 16

**16-1** Page 549, Section 16.2.2.2, Eq.(16.9)

It should be  $y_i \neq \hat{y}_c$

**16-2** Page 550, Section 16.2.2.2, under Eq.(16.14)

$(1 - \hat{\pi}_c$  should be  $(1 - \hat{\pi}_c)$ .

### XVII. CHAP 17

**17-1** Page 599, Section 17.2.3.1, in Eq. (17.23)

The first term  $\pi_1(1 - A_{12} - A_{12})$  should be  $\pi_1(1 - A_{12} - A_{13})$ .

**17-2** Page 599, Section 17.2.3.2, the 2nd paragraph

$\pi 1_{K \times 1} = 0$  should be  $\pi 1_{K \times 1} = 1$ .

**17-3** Page 614, Section 17.4.3.2, in Eq. (17.66)

$\phi_{t+1}(j)\beta_{t+1}(j)\psi(i, j)$  should be  $\psi_{t+1}(j)\beta_{t+1}(j)\Psi(i, j)$ .

Please refer to the paragraph below Eq.(17.48) for their definitions.

**17-4** Page 614, Section 17.4.3.2, in Eq. (17.67)

$\phi_{t+1}$  should be  $\psi_{t+1}$ .

**17-5** Page 618 , above Section 17.4.4.5, in Fig 17.13

Third line: 'Each state  $z_1, z_2, z_3$ ' should be 'Each state  $s_1, s_2, s_3$ '

## XVIII. CHAP 18

**18-1** Page 635, Section 18.2.1, under Eq.(18.9)

$\mathcal{N}(0, Q)$  should be  $\mathcal{N}(0, Q_t)$ .

**18-2** Page 635, Section 18.2.1, in Eq.(18.11)

The last term,

$$\begin{pmatrix} \delta_{1t} \\ \delta_{2t} \\ \delta_{3t} \\ \delta_{4t} \end{pmatrix}$$

should be

$$\begin{pmatrix} \delta_{1t} \\ \delta_{2t} \end{pmatrix}$$

**18-3** Page 635, Section 18.2.1, under Eq.(18.11)

$\mathcal{N}(0, R)$  should be  $\mathcal{N}(0, R_t)$ .

**18-4** Page 640, Section 18.2.4.2, under Eq. (18.21)

$a_t = a_{t-1} + b_0 t$  should be  $a_t = a_{t-1} + b_0$ .

**18-5** Page 643, Section 18.3.1.2, in Eq. (18.30)

$p(z_t|y_{1:t}, u_t)$  should be  $p(z_t|y_{1:t}, u_{1:t})$ .

**18-6** Page 643, Section 18.3.1.2, in Eq. (18.39)

Subscripts are missed. It should be

$$K_t = \Sigma_{t|t-1} C_t^T (C_t \Sigma_{t|t-1} C_t^T + R_t)^{-1} = (\Sigma_{t|t-1}^{-1} + C_t^T R_t C_t)^{-1} C_t^T R_t^{-1}$$

**18-7** Page 644, Section 18.3.1.4, in Eqs. (18.42-43)

$C$  should be  $C_t$ .

## **XIX. CHAP 19**

**19-1** Page 671, Section 19.4.1, following Eq. (19.19)

It says 'where  $\theta = (W, b)$ '. But there's no  $\theta$ . Maybe the probability appeared around can be rewritten as conditioned on  $\theta$ .

**19-2** Page 674, Section 19.4.4, in Eq. (19.28)

An explanation for  $y_{-(st)}$  is needed here.

**19-3** Page 680, Section 19.5.2, under Eq.(19.46)

In the 1st line below Eq.(19.46),  $\sum_{h_i} \tilde{p}(y_i, h_i|\theta)$  should be  $\sum_h \tilde{p}(y_i, h|\theta)$ , to be consistent.

**19-4** Page 682, Section 19.5.5, Algorithm 19.1

It's better to use the same notations as in Eq.(19.51):  $B$  to be  $N$ ,  $\hat{E}$  to be  $\mathbb{E}$ .

## **XX. CHAP 20**

**20-1** Page 710, Section 20.2.1, above Eq. (20.2)

$m_{u \rightarrow t}(x_t)$  should be  $m_{u \rightarrow t}^-(x_t)$ .

**20-2** Page 715, Section 20.2.4.3, under Eq.(20.37)

2nd line following Eq.(20.37):  $\phi_{t+1}$  should be  $\psi_{t+1}$ .

**20-3** Page 720, Section 20.3.2, Table 20.3

In the 1st line,  $\psi_L(L, )$  should be  $\psi_L(L, G)$ .

**20-4** Page 727, Section 20.4.2.1, in Eq. (20.69)

A ')' is left on the rightmost.

## **XXI. CHAP 21**

**21-1** Page 738, Section 21.3, in Eq. (21.26)

$q(x_1)$  should be  $q_1(x_1)$ .

**21-2** Page 740, Section 21.3.2, in Eq. (21.41)

The  $-$  should be  $+$ .



**21-3** Page 761, Section 21.8.2.1, in Eqs. (21.168-170)

$p(w|\mathcal{D})$  should be  $p(w)$ .

**21-4** Page 762, Section 21.8.2.1, in Eq. (21.176) and Eq. (21.179)

$V_n$  should be  $V_N$ .

**21-5** Page 762, Section 21.8.2.1, in Algorithm 21.1, line 8

$\text{reshape}(m, M, N)$  should be  $\text{reshape}(\psi, M, N)$ .

## XXII. CHAP 22

**22-1** Page 784, Section 22.3.5.4, under Eq. (22.44)

1st following Eq. (22.44),  $\sum_{x_t} \tau(x_s, x_t)$  should be  $\sum_{x_t} \tau_{st}(x_s, x_t)$ .

**22-2** Page 803, Section 22.6.3.1, in Eq. (22.151)

$\lambda_{st}$  should be  $\lambda_{u,v}$  to be consistent with the followings.

## XXIII. CHAP 23

Not found.

## XXIV. CHAP 24

**24-1** Page 841, Section 24.2.2, in Eq.(24.2)

In the denominator,  $s_t$  should be  $x_t$ .

**24-2** Page 841, Section 24.2.2, under Eq.(24.4)

$x_t$  should be  $x_s$ .

**24-3** Page 871, Section 24.6, in Figure (24.17) b

$T = 0.5$  should be  $T = 0.2$ .

## XXV. CHAP 25

**25-1** Page 883, Section 25.2, Figure 25.3

Figure 25.3 is chaotic. It seems the six subplots should be arranged to be a  $2 \times 3$  matrix, instead of being a  $3 \times 2$ .

## XXVI. CHAP 26

Not found.

## XXVII. CHAP 27

**27-1** Page 959, Section 27.3.4, in Eq. (27.31)

$(z_{il} = k)$  should be  $(q_{il} = k)$ .

**27-2** Page 959, Section 27.3.4, in Eq. (27.32)

$(x_{il} = v, z_{il} = k)$  should be  $(y_{il} = v, q_{il} = k)$ .

**27-3** Page 960, Section

**27-4** Page , Section 27.3.4, in Eq. (27.36)

$\beta$  should be  $\gamma$ .

**27-5** Page 990, Section 27.7.1.2, in Eq. (27.107)

$P(h_k = 1|c, \theta)$  should be  $P(h_k = 1|v, \theta)$ .

**27-6** Page 992, Section 27.7.2.2, in Eq. (27.116)

On the rightmost,  $h_r$  should be  $h_k$ .

## XXVIII. CHAP 28

Not found.