Book Errata and Updates for "Statistical Machine Learning: A unified framework" by Richard M. Golden

Richard Golden

March 31, 2023

Page 3. (2nd line from bottom of page)

Replace
The process which generates the training data is called the *data generating process*.
with
The process which generates the data is called the *data generating process*.

Page 17. Example 1.3.6.

Replace

$$\hat{\ell}_n(\boldsymbol{\theta}) = (1/n)\sum_{i=1}^{n}(y_i - \ddot{y}(\mathbf{s}_i, \boldsymbol{\theta})^2$$

with

$$\hat{\ell}_n(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \ddot{y}(\mathbf{s}_i, \boldsymbol{\theta})^2$$

Page 43. Second Paragraph

Replace
A *cluster* C_k is a subset of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$
with
A *cluster* C_k is a subset of $\{1, \ldots, n\}$ defined such that $j \in C_k$ means that $\mathbf{x}_j$ belongs to cluster C_k.

Page 46. Example 1.6.5

Replace

$$V(\mathbf{y}_1, \ldots, \mathbf{y}_n) = -\sum_{i=1}^{n}\sum_{j\neq i} p(\mathbf{x}_i|\mathbf{x}_j)\log q(\mathbf{y}_i|\mathbf{y}_j)$$

with

$$V(\mathbf{y}_1, \ldots, \mathbf{y}_n) = -\sum_{i=1}^{n}\sum_{j\neq i} p_x(\mathbf{x}_i|\mathbf{x}_j)\log p_y(\mathbf{y}_i|\mathbf{y}_j)$$

Page 127. Example 5.1.11.

Replace

$$\text{where } \phi(\psi) = \psi$$

with

$$\text{where } \phi(\psi) = \psi + 1$$

And ALSO Replace

$$\ddot{\phi}(\psi) = \tau \log(1 + \exp(\psi/\tau))$$

with

$$\ddot{\phi}(\psi) = \tau \log(1 + \exp((\psi + 1)/\tau))$$

Page 139. Bottom of Page

Replace
bfu_{d_j}
with
$\mathbf{u}_{d_j}$

Page 140. Problem 5.2-1.

Replace
Show that $d\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right)/d\mathbf{x} = \left[\mathbf{B} + \mathbf{B}^T\right]\mathbf{x}$
with
Show that $d\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right)/d\mathbf{x} = \mathbf{x}^T\left[\mathbf{B} + \mathbf{B}^T\right]$

Page 140. Problem 5.2-2.

Replace
Show that $d\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right)/d\mathbf{x} = 2\mathbf{B}\mathbf{x}$ where $\mathbf{B}$ is a symmetric matrix.
with
Show that $d\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right)/d\mathbf{x} = 2\mathbf{x}^T\mathbf{B}$ where $\mathbf{B}$ is a symmetric matrix.

Page 158. Equation 5.45

The first term on the right-hand side of the equation should be negative and not positive. The second term on the right-hand side of the equation should be positive and not negative. Compare with the definition of the Lagrangian in the equation on the previous line.

Page 161. Equation 5.57

Derivative in Equation (5.57) on the left-hand side should be with respect to r.

Page 161. Equation 5.59

Derivative in Equation (5.59) on the left-hand side should be with respect to $\mathbf{h}$.

Page 195. Problem 7.1-5.

The notation $dV_{x,d}(\gamma_{max})$ which occurs twice in the description of this problem should be replaced (twice) with the notation $dV_{x,d}(\gamma_{max})/d\gamma$.

Page 197

Equations (7.30), (7.31), and (7.32) contain summations which range from $t = 0$ to $t = M$. We want to modify these three equations so that the summations rangeo from $t = 0$ to $t = M - 1$.

Page 204

Replace all the terms in the four equations which have the form $O(\gamma(t)^2)$ with $O(\gamma(t)^2)\mathbf{1}_d$.

Page 255. Definition 9.3.6.

Replace Equation (9.6) given by:

$$p(\mathbf{x};\boldsymbol{\theta}) = (2\pi)^{-d/2}(\det(\mathbf{C}))^{-1/2}\exp\left[-(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

with

$$p(\mathbf{x};\boldsymbol{\theta}) = (2\pi)^{-d/2}(\det(\mathbf{C}))^{-1/2}\exp\left[-(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})/2\right]$$

Page 273. Definition 10.2.1.1

The energy function V includes a negative sign which should be omitted. The revised version should really write out the probability density functions explicitly for all of the examples/definitions in Section 10.2.

Page 318

[equation abouit 16 lines from bottom of page] Right above the sentence which begins "Therefore," insert to the right of the equality sign on the left $\exp[-V(\mathbf{c}^k) + V(\mathbf{x})]$

Page 352, Equation 12.55

The "minus" sign in the equation should be a "plus" sign.

Statement and Proof of Theorem 13.1.2 (Page 366)

Replace
Let Θ be a closed, bounded, and convex subset of $\mathcal{R}^q$ that contains a local minimizer, $\boldsymbol{\theta}^*$,...
With
Let Θ be a closed, bounded, and convex subset of $\mathcal{R}^q$ whose interior contains the local minimizer $\boldsymbol{\theta}^*$...

Page 367, Recipe Box 13.1, Step 2

Replace the phrase "continuous random" which appears twice with "continuous". Alternatively, you can replace "continuous random" with "continuous random function which is piecewise continuous in its first argument".

Page 395, 4 lines below Equation 3.61

Replace $p(\eta|\mathcal{D}_n), \bar{\eta})$ with $p(\eta|\mathcal{D}_n), \bar{\eta}$.

Page 407, Definition 14.2.2

The definition of $\hat{p}_e(\mathcal{D}_n) = \prod_{i=1}^{n} \hat{p}(\mathbf{x}_i)$ should be: $\hat{p}_e(\mathcal{D}_n) = \prod_{i=1}^{n} \hat{p}_e(\mathbf{x}_i)$

Page 420. Last Paragraph Before Example 15.11

Insert sentence: "The matrix $\hat{\mathbf{B}}_n$ is an approximation for $\mathbf{B}^*$." after first sentence in this paragraph. Then replace "but not a sufficient condition for **A8** to hold" with "but not a sufficient condition for $\hat{\mathbf{B}}_n$ to be positive definite is that the number of training stimuli is greater than or equal to the number of free parameters."

Page 427. Recipe Box

The notation $\mathbf{g}_i(\boldsymbol{\theta})$ should be used to replace $\hat{g}_i$.

Page 429. Recipe Box.

Missing inverse on $\hat{\mathbf{A}}_n$ should be $\hat{\mathbf{C}}_n = (\hat{\mathbf{A}}_n)^{-1}\hat{\mathbf{B}}_n(\hat{\mathbf{A}}_n)^{-1}$.

Page 451. Equation 16.5

There is a missing "trace" operator on the right-hand side of Equation 16-4. The left-hand side is a scalar and the right-hand side is a matrix.