A First Course in Machine Learning – errata

Simon Rogers Mark Girolami

February 19, 2016

- p9, para1, $\widehat{x_1} \rightarrow \widehat{w_1}$
- p18, Comment 1.7, the second row of the matrix multiplication example is erroneously a copy of the first. The final matrix should be:

$$\begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} \end{bmatrix}$$

- p15, section1.3, $f(x, s_1, \ldots, s_5; w_0, \ldots, w_6) \rightarrow f(x, s_1, \ldots, s_8; w_0, \ldots, w_9)$
- p52, Equation 2.13 and the equation block below it should read (an X had been replaced with an x!):

$$\mathsf{var}\{X\} = \mathbf{E}_{p(x)}\left\{(X - \mathbf{E}_{p(x)}\{X\})^2\right\}$$

$$\begin{aligned} \mathsf{var}\{X\} &= \mathbf{E}_{p(x)}\left\{(X - \mathbf{E}_{p(x)}\{X\})^2\right\} \\ &= \mathbf{E}_{p(x)}\left\{X^2 - 2X\mathbf{E}_{p(x)}\{X\} + \mathbf{E}_{p(x)}\{X\}^2\right\} \\ &= \mathbf{E}_{p(x)}\left\{X^2\right\} - 2\mathbf{E}_{p(x)}\{X\}\,\mathbf{E}_{p(x)}\{X\} + \mathbf{E}_{p(x)}\{X\}^2 \end{aligned}$$

- p70, Equation 2.34. The Left-Hand-Side should be:

$$\frac{\delta \log L}{\delta \sigma}$$

- p34, second line, reguarised →regularised
- p60, first equation, $=\rightarrow\approx$
- p79, equation 2.44, $\mathbf{E}_{p(\mathbf{t}|\mathbf{X},\mathbf{w},\sigma^2)}\left\{\mathbf{w}\mathbf{w}^\mathsf{T}\right\}$ should be $\mathbf{E}_{p(\mathbf{t}|\mathbf{X},\mathbf{w},\sigma^2)}\left\{\widehat{\mathbf{w}}\widehat{\mathbf{w}}^\mathsf{T}\right\}$
- p80, . . . is the negative inverse of something called the Fisher information matrix →. . . is the inverse of something called the Fisher information matrix
- p98, para2, This is much lower than that for $r = 5$→This is much lower than that for $r = 0.5$.
- p103, para1, . . . will gave a. . . →will give a
- p103, equation 3.6. The power of the $(1 - r)$ term is missing a -1:

$$p(r|y_N) = \frac{\Gamma(\alpha + \beta + N)}{\Gamma(\alpha + y_N)\Gamma(\beta + N - y_N)} r^{\alpha + y_N - 1}(1 - r)^{\beta + N - y_N - 1}$$

- p123, below Equation 3-16. '. . . doesn't depend on $\mathbf{w}_{\mathsf{new}}$ and so. . . ' →'. . . doesn't depend on $\mathbf{x}_{\mathsf{new}}$ and so. . . '
- p130, para3, . . . Gaussian density ($\mathcal{N}(\mathbf{x}_{\mathsf{new}}^\mathsf{T}\mathbf{w}, \sigma^2)$ is. . . →($\mathcal{N}(\mathbf{x}_{\mathsf{new}}^\mathsf{T}\mathbf{w}, \sigma^2)$)
- p134, The hint for exercise EX 3.5 has its dr in the wrong place. The equation should be:

$$\int_{r=0}^{r=1} r^{a-1}(1 - r)^{b-1}\, dr = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$$

- p137, final reference, mulitple →multiple . . . Gaussian density ($\mathcal{N}(\mathbf{x}_{\mathsf{new}}^\mathsf{T}\mathbf{w}, \sigma^2)$) is. . .

- p137, ref[1], mcmc→MCMC
- p143, there is an issue with the normalising constant. As:

$$Z^{-1} = p(\mathbf{t}|\mathbf{X}, \sigma^2)$$

  then the posterior should be:

$$p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \sigma^2) = Zg(\mathbf{w}; \mathbf{X}, \mathbf{t}, \sigma^2).$$

- p145, In the second line of the first set of equations, the log operates on both fractions. I.e.:

$$= \sum_{n=1}^{N} \log \left[\left(\frac{1}{1 + \exp(-\mathbf{w}^T \mathbf{x}_n)} \right)^{t_n} \left(\frac{\exp(-\mathbf{w}^T \mathbf{x}_n)}{1 + \exp(-\mathbf{w}^T \mathbf{x}_n)} \right)^{1-t_n} \right]$$

- p168, ref[10], Alex Smola was an editor not a co-author of this paper.
- p141 , para2, prior density, $p(\mathbf{w})$ exercises.→prior density, $p(\mathbf{w})$.
- p154, para4, the printing process seems to have removed the white areas from the dartboard. These should be the two large areas in each segment.
- p188, in comment 5.1, the 'b' in the penultimate line should be an 'a'.
- p195, in the SVM objective function, the second summation should be over n and m:

$$\underset{\mathbf{w}}{argmax} \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n,m=1}^{N} \alpha_n \alpha_m t_n t_m k(\mathbf{x}_n, \mathbf{x}_m)$$

- p202, In the table, there are two columns with the heading '18' – the first should be '17'.
- p233, last paragraph, However, it is comes with →However, it comes with
- p242, third paragraph, ... must be orthogonal to $\mathbf{w}_1$ ($\mathbf{w}_1^T \mathbf{x}_2 = 0$) →must be orthogonal to $\mathbf{w}_1$ ($\mathbf{w}_1^T \mathbf{w}_2 = 0$)
- p244, comment7.1 eigen values→eigenvalues
- p255, Equation block 7.11 is formatted badly and misses a db it should be:

$$\begin{aligned} \mathbf{E}_{p(a)p(b)}\{f(a)f(b)\} &= \iint p(a)p(b)f(a)f(b)\ da\ db \\ &= \iint p(a)f(a)\ da\ p(b)f(b)\ db \\ &= \int \mathbf{E}_{p(a)}\{f(a)\}p(b)f(b)\ db \\ &= \mathbf{E}_{p(a)}\{f(a)\}\mathbf{E}_{p(b)}\{f(b)\} \end{aligned}$$

- p255, $\mathbf{E}_{Q_{\mathbf{x}_n}(\mathbf{x}_n)Q_{\mathbf{w}_m}(\mathbf{w}_m)}\left\{\mathbf{x}_m^T \mathbf{w}_m \mathbf{w}_m^T \mathbf{x}_n\right\}$→$\mathbf{E}_{Q_{\mathbf{x}_n}(\mathbf{x}_n)Q_{\mathbf{w}_m}(\mathbf{w}_m)}\left\{\mathbf{x}_n^T \mathbf{w}_m \mathbf{w}_m^T \mathbf{x}_n\right\}$

Acknowledgements: Hat-tip to the observant individuals who have found these: Gavin Cawley, Luisa Pinto, Andy O'Hanley, Joanathan Balkind, Mark Mitterdorfer, Charalampos Chanialidis, Juan F. Saa, Eu Wern Teh, Qin Zhou, ...