Errata in “Machine learning: a probabilistic perspective”

Below are known errors in the first printing, as well as small edits that I have made, including new citations. These changes will appear in the third printing.

Pages refer to first print edition (6/20/12)

- Long list of errors found by Sebastian Bratieres can be found here https://docs.google.com/document/pub?id=10MAgk3R2GhzoDPG3lv5vODBTBNQlZxygLmkmv2MTWPE

-p4 Added reference to Nate Silver book (‘The Signal and the Noise’) and reformatted sec 1.2.1.2

- footnote 3 on p31: 'risk reverse reward' should be 'risk versus reward'

- p35 After eqn 2.35: ref to “Figure 2.1(b-c)” should say “Figure 2.1”.

- sec 2.4.2 on student T distribution was missing subsection title; this change has caused all subsequent subsections in 2.4 to be renumbered

- p.68: iff (if and only if) [and not "iff (***iff*** and only if)]

- equation 3.14 (p.74). should be
p(theta | D) equiv. p(D | theta) * p(theta) = (...)

- p.74, 2nd parag.: missing ')' after first theta in p(D | theta***)*** equiv. P(s(D) | theta)

- p.74, end of 1st line in section 3.3.2: missing 'be' in "it would ***be*** convenient (...)"

- p.74, penultimate parag.: "and that we think it lives in the interval (0.05, 0.30) with
probability ***???***, then ..." [is a value missing there???]

- Page 84, line 8 of Algorithm 3.1. for theta_jc, you should have N_c, not N.

- page 85, equation 3.67, the subindex k should be c.

- ex 4.8c. Whitening uses U and Lambda, which are the eigenvectors / values of X'X, not X

- eqn 6.49 should be R(\data, \delta)

-p250 top paragraph. “If not, the objective function is not convex” -> “If the objective function is not convex”

- p234 Added footnote about Nate Silver flood prediction example

- sec 8.3.1 Defined mu_i = sigmoid(w’ x_i) and fixed typo p(tilde{y}=-1) = 1/(1+exp(w’x))

- Eqns 8.53-8.54 are wrong (confusion of p(theta,data) and p(theta|data)).
The correct equations are as follows:

```
\bea
\hat{p}(\vtheta,\data) &=&  e^{-E(\vtheta^*)}
\exp\left[ -\frac{1}{2}   (\vtheta-\vtheta^*)^T \vH (\vtheta-\vtheta^*) \right]
\\
\hat{p}(\vtheta|\data) &=&  \frac{1}{Z} \hat{p}(\vtheta,\data)
\propto \gauss(\vtheta|\vtheta^*,\vH^{-1})
\\
Z &=& e^{-E(\vtheta^*)}  (2\pi)^{\Ndim/2} |\vH|^{-\half}
\eea
```

- Eq. 8.65 E[a^2] -> E[a]^2

- Eq. 8.68, 8.69, 8.70.. \mu \sigma -> \mu_a \sigma_a There are other places where you drop the subscript a.

- Eq. 8.68 \Phi(\frac{a}{\lambda^{-2} + ...}) I guess this should be \Phi(\frac{\mu_a}{...})

- p264 Added a new paragraph to the end of 8.5.2.3

- sec 12.5.1. Forward references to CCA should be removed

- p440 bottom paragraph. “5.108, which showed that that” -> “5.108, which showed that”

- p443 ||theta - y|_2^2 -> ||theta - y||_2^2

- p445 epxanding -> expanding

- p445 New footnote added about mirror descent

- p446 Modified caption of algo 13.2

- Sec 14.3.2 penultimate sentence should say “RVM is the sparsest, then L1VM, then SVM”.

- Caption of fig 14.6 should say “for the models in fig 14.5”

- algo 17.1 line 5 should be log p(x_{1:T})

- sec 17.6.1.1 p622 Last sentence should say “Thereafter”

- eqn 19.49 should be p_emp(y)

- eqn 19.50 should be p_emp(y), no sum over x

- eqn 19.79 should be L(y, yhat)

- eqn 19.83 should be Z(x_i, w)

- eqn 19.84 vy_i should be \vy_i

- exercise 22.3 on p812 is now credited to Boykov and Veksler

- eq 24.110 Final integral should be int p(theta) d theta,
not p(theta|D)

- fig 27.29a is incorrect (shows fig 2b instead of 2a from Salakhudinov’08)

- eqn 27.94 is missing the sigmoid term. It should be
p(R(u,v)|\vx_u, \vx_v,\vtheta) = Ber(\sigmoid((\vU \vx_u)^T (\vV \vx_v)))

- p1006 Added reference to Dean’12 (Google Brain project)

Other Errors, found by Robert Piche. Page numbers refer to 4/23/12 pre-publication edition. These will be fixed in the 3rd printing.

p.14 masket basket -> market basket

p. 29 risk reverse -> risk averse

p. 30: This [cdf] is obviously a monotonically increasing function.
[no, but it is a monotonically non-decreasing function]

p. 31 rightost cutoff -> rightmost cutoff

p.31 Since the cdf F is a monotonically increasing function, it has an inverse;
[Some cdf's are not monotonically increasing and so do not have an inverse]

p. 31 contain alpha/2 probability mass -> contain alpha/2 of the probability mass

p.31 If this integral is not finite, the mean is not defined
[should be: if \int |x| p(x) dx is not finite then the mean is not defined.]

p.34 column 7 is all G's -> column 13 is all G's

p. 34 Figure 2.5 (b) does not agree with the data in (a), e.g. column 1 doesn't contain the letter c at all!

p.59 betweeb -> between

p.60 and many other occurrences: Bayes rule -> Bayes' rule or Bayes's rule or The Bayes rule

p.61 there exist function -> there exist functions

p.75 likelihood: since -> likelihood, since

p. 76 is convex -> is a convex

p.92 dirichlet -> Dirichlet

p. 95 bayes spam -> Bayes spam

p. 98 elogonated -> elongated

p. 101 addtional -> additional

p. 101 the the -> the

p. 104 is so-called -> is so called

p. 146 the mode is a point of measure zero -> [redundant: all points have measure zero.
Rewrite as, e.g.: the mode is determined by a single point of the density]

p. 158 parametrs -> parameters

p.160 In Figure 5.9(b) shows -> Figure 5.9(b) shows

p. 168 Figure 5.11 a [The last two plots are meant to be compared and so should have the same y-axis range.]

p.169 example, example -> example, but in this example

p.175 where L_{FN} -> [remove indentation]

p.181 absestos -> asbestos [two occurrences]

p.213 esitmate -> estimate

p.219 \frac{d}{d \theta} -> \frac{d^2}{d \theta^2}

p.220 is C_1 continuous -> has continuous first derivative [the C_1 notation hasn't been introduced yet]

p.223 In Section 1.4.8, we will discuss a more probabilistic approach -> [wrong Section pointer?]

p.224 sommon -> common

p.232 Student T -> Student t

p.232 This latter terms -> This latter term

p.232 Eqn 7.82 mismatched parenthesis

p.234 marignal -> marginal

p.234 be the precision -> is the precision

p.234 principle practical advantage -> principal practical advantage

p.248 Eqn (8.30): \lambda w -> 2 \lambda w

p.249 kronecker -> Kronecker

p.251 bu the likelihood -> by the likelihood

p. 251 a a Gaussian -> a Gaussian

p.252 Figures 8.5c and d should have a grid too

p.253 results in a highly skewed posterior -> ? [Figure 8.5c contours look the same as those in

8.5b]

p.205 & 253: different spellings of focuses / focusses

p.253 posteiror -> posterior

p.259 one of the main drawback -> one of the main drawbacks

p. 260 see Algorithm 8 -> see Algorithm 8.3

p.262 see Algorithm 11 -> see Algorithm 8.4

p. 264 Salojarvi et al. (2005) -> (Salojarvi et al. 2005)

p.264 Little. and Rubin -> Little and Rubin

p. 267 soon we will study are -> soon we will study

p. 269 that that -> that

p. 271 this this -> this