Changes to MLAPA that will appear in the fourth printing (to appear September 2013)

Note that page numbers change across different print editions, but the section numbers, figure numbers and equation numbers are constant.

- Fixed various trivial typos (eg using - instead of -- in the bibliography) found by MIT press.

- **Fixed the errors found by Sebastian Brattieres and friends (up to 8/11/13), listed here:**
https://docs.google.com/document/d/10MAgk3R2GhzoDPG3lv5vODBTBNQlZxygLmkmv2MTWPE/edit#heading=h.ywbhd3xp4bnk

- **Fixed the errors found by readers (up to 8/11/13) listed here:**
https://docs.google.com/spreadsheet/ccc?key=0AvhicYJxvbiOdEp2U2hRWXVpMU5nd05YcEJKVFNUdmc#gid=0

- preface: added printing history to the end of the preface, to make it easier for the reader to determine which version they have

- sec 1.1: reworded description of the long tail

- sec 2.2.3 Added quote by Jeffreys'73 and ref to the book by Sharon McGrayne (2011) to the Bayes rule section

- eqn 3.90 (in exercise 3.3) typo: should be Gamma(a0+a1+1) = (a0+a1) Gamma(a0+a1)

- eqn 3.93 (exercise 3.9) should be max(D, b) instead of max(D)

- eqn 4.62 Should be delta_c = x' beta_c + gamma_c

- eqn 4.82 Should be \vmu_{1|2} = -\vL_1^{-1} \vL_2 \vx_2

- eqn 4.181 Should be SN = SO + Smu, not inv(SN)

- eqn 4.197 Should be exp(-1/2 tr(Sigma(-1) S)) not exp(-N/2 …)

- sec 4.5 Distribution of scatter matrix is S ~ Wi(Sigma, N) not Wi(Sigma, 1)

- sec 5.7.2.1 Added footnote that AUC is equivalent to a pairwise ranking criterion

- eqn 7.27 should read

min sum_i (r_i^+) + (r_i^-) st r_i^+>0, r_i^- >0, w'x_i +(r_i^+) - (r_i^-) = y_i

- above eqn 7.44, should be $Z = U S$ not $Z = U D$

- sec 8.1-8.2: 'Intro to Logistic regression' has been expanded, to provide more context and background (as suggested by Jonni Kanerva).

- eqn 8.9 is actually not a convex function, although the Hessian is positive definite in the vicinity of the global minimum

- eqn 9.18 Denominator of last term missing +1

- above 9.100 Should be $z_i = u_{1i} - u_{0i}$ without $epsilon_i$

- sec 9.7.1. A few tweaks to the wording.

- sec 9.7.2 Added ref to Rendle09 (Bayesian Personalized Ranking paper)

- eqn 10.22 We see that U (but not W) is part of the Cholesky decomp of Sigma

- sec 10.5.1 Example of d-separation is all wrong! We need to condition on x_1 to break the 2->6 path via x1.

- sec 11.4.8 p365 Typo in Liang and Klein citation

- sec 12.1.13 Added footnote about identifiability for Bayesian confirmatory FA

- sec 12.2.1 Added footnote about FA being invariant to rescaling of data (unlike PCA)

- sec 12.2.4 minor typo on page 396 line 4: "if sigma^2 >, ..." is missing a 0.

- Thm 12.2.2. No need for W to be orthogonal for the theorem to hold.

- sec 12.4 Added mention of item response theory and connection to categorical PCA

- eqn 12.5 missing $E[z z']$ term

- fig 14.4d Sebastien noticed that all the circles were only on orange points. This was due to a bug in svmlightFit.m line 87. It should be
  model.svi = find(abs(alpha) > epsilon);
Rerunning the code now shows that 196 out of the 200 data points become support vectors (this is with C chosen by cross validation, as before). New figure is in book.

- fig 14.5d Same problem with the support vectors as above. Now fixed. Note: we set C = 1/lambda, to be consistent with other plots. New figure is in book, matches new 14.6d.

- sec 15.3 Added ref to Chen and Dong 2011 (Generalized GPs)

- sec 16.5 Added ref to ReLU

- sec 16.5.2. Added ref to LSTM paper and Mikolov'11.

- fig 17.13 The Emission distribution for state 1 should be [C1: 0.5, c2: 0.2, c3: 0.3] for the example on the RHS to be correct, not [0.5, 0.3, 0.2]

- eqn 17.45 Added footnote to clarify distinction between $alpha\_t(j) = Pr(z\_t=j \mid x(1:t))$ and the more common definition, $alpha\_t(j) = Pr(z\_t=j, x(1:t))$

- eqn 18.96 and 18.97: LHS should be w_m^0 and w_c^0 not w_m^i and w_c^i

- sec 19.2.1 Markov blanket of 5 is {2,3,4,6,7}

- sec 19.4.4.1 Sparse directed Gaussian DGMs do not correspond to sparse cholesky decompositions of Sigma... (see sec 10.2.5 for the correct statement)

- sec 19.5.3. (MRF training) Added ref to Kulesza and Pereira 2007 on dangers of using approx inf inside of MRF/ CRF learning

- sec 19.6.3 (CRF training) Added ref to Wick11 (Sample Rank), and included a brief description. Also added more cross references to other parts of the book that discuss MRF learning methods.

- exercise 20.1 referred to the wrong figure; new image has been added

- exercise 21.1 and 21.2 are duplicates

- eqn 21.20 should be \be
q(\vx) = \gauss(x_1|\mu_1,\Sigma_{11}) \gauss(x_2|\mu_2,\Sigma_{22})
where $\vSigma=\vLambda^{-1}$.

- eqn 21.30 Last q_j term should be $log(q\_j)$

- eqn 21.39 RHS should be $exp(sum\_i -L\_i(x\_i))$

- sec 21.7 Added ref to Knowles and Minka 2011

- sec 22.6.4 Added ref to Kappes et al 2013

- eqn 22.130 is wrong. The y_g term should pre-multiply the sigma*Psi term, not the mu.

- eqn 22.134 Denominator should be {m^{t}_{h_g \ra d_g}(d_g)}

- sec 28.3.1 Added more refs on training methods for deep neural nets

Typos found by Olcan Sercinoglu in first printing
page 246, section 8.3.1, just before eq. 8.4, one of the +1s should be -1.
in equations 8.34 through 8.40 there may be some confusion about big W and small w - fixed.

p 381 - Gaussian spray can: prior is isotropic, mu should be bold.

page 396 line 4: "if sigma^2 >, ..." is missing a 0.

Page 463, first paragraph of 13.7.1, "case can be"

Page 555, last equation (16.26), second equals sign

Page 557, equation 16.31, minus sign

Page 594, equation 17.16 alpha should be alpha(j)

Typos found by Jutta Degener in first printing

p63 exercise 2.11 "a ID" -> "an ID"

below 2.135: Evaluate the integral and hence show Z = sigma root(two pi).
The parentheses around two pi were unexpected, shouldn't that just
be typeset under the root's arm?

p64 Exercise 2.17 left most -> leftmost

p70 Below Figure 3.3, "MAP/ ML" -> "MAP/ML"

p72 above 3.2.5, "see**ms**" -> "sees"

p78 just above 3.4: "blackswan type" -> "black-swan type"

3.4 "a di**ce**" -> "a die".

p81 "or the ing from verbs" -- but then the example you quote also
repairs a consonant doubling: "running" -> "run". Maybe "raining" -> "rain"
would sidestep that distraction?

The correct encoding of the second row of "Mary had a little lamb"

with stopwords removed is

    1 10 3 2 10 5 6 8

the original has 1 10 3 2 10 5 **10** 6 8, encoding the explicitly
listed stopword "as" falsely as "unk".

p82 "class conditional density" -> "class-conditional density"
"one dimensional densities" -> "one-dimensional densities".

p83 "we can **model** use" -> "we can use"

p87 above 3.5.5, "**Figure** 3.1" -> "Table 3.1"

p92 Exercise 3.13, "**d**irichlet" -> "Dirichlet"

p93 Hint below 3.104: "Hint: recall that**,** for a vector of counts**,** N1:K, the"
should be: "Hint: recall that for a vector of counts N1:K, the"

p94 below 3.108: "succ**int**ly" -> "succinctly"

p95 The formatting in exercise 3.22 is thrown off by the vertical bars,
which render much taller than the plain "spam" in the first theta.
(I.e. all but the first spam get lots of vertical space.)

p101 "add**ti**onal" -> "additional"

p106 4.2.3 "Furthemore" -> "Furthermore" (add r)

   "half way" -> "halfway"

p107 "Use a diagonal covariance matrix _and_ forced it to be shared."
    forced -> force

  "low dimensional subspace" -> "low-dimensional subspace"

  in 4.2.6, "lower dimensional space" -> "lower-dimensional space"

p108 "In high dimensional settings" -> "In high-dimensional settings"

p109 4.2.8 "In high dimensional problems" -> "In high-dimensional problems"

p112 4.3.2.1, I don't understand why mu = 0 is typeset in bold.

p114 4.3.2.3 "a 20 dimensional Gaussian" -> "a 20-dimensional Gaussian"

p115 4.3.3 "where we use the notation Nc() to distinguish from the

moment parameterization N()."
Although it's clear what you mean, this would be easier to read
for me if "distinguish" had an object. E.g., "it".

p118 4.3.4.2 "The first two equations are s known"
-> .. are also known ...

"in machine learning/ statistics"
fix asymmetric spacing around "/" .

p120 4.4.2.1 "Bayes rule" -> "Bayes' rule"
(and many more times throughout)

p122 4.4.2.2 "is half way between" -> "is halfway between"

p130 4.6.2.1 "This begs the question:"
Nobody will misunderstand this, but it'll distract a few pedants.
Did you know that that phrase has drifted in meaning from its
original sense, "this argument is circular by asking the reader
to assume its conclusion", to "This invites the question:" ?
I can see why few people care for the original meaning -- it's kind of
convoluted -- but since it's such a pleasantly old-fashioned phrase,
I feel a little bad for it, since we're stealing the old-fashioned
sound without the old-fashioned meaning.
So, in the interest of stylistic honesty, s/begs/invites/.

"the off diagonal elements" -> "the off-diagonal elements"

"a 50 dimensional Gaussian" -> "a 50-dimensional Gaussian"

p133 "Normal-inverse-wishart" -> "Normal-inverse-Wishart", surely?

p134 "Student T" -> "Student's t" (throughout)

p135 "Student-T distribution" -> "Student's t distribution"

"xi is 1d" -> "xi is 1D" or 1-D (just pick something and stick with it),
throughout

p136 "of the student" -> "of Student's distribution"
(Or stick with lower-case throughout; you could make an argument
for that, since it really is not a person's name--but I think
Student's t is the dominant form at this point)

p140 "over confidence" -> "overconfidence"

p141 Equation 4.268 is too long and overlapping its tag.

p142 Exercise 4.8a "where have labeled" -> "where we have labeled"

p144 "How big does n have to be ... is the data." -> end with a ?, since
    it's a question

    Exercise 4.15: "The unbiased estimates ... is given by"
      Number agreement: either make it one estimate, or "are given by".

p145 "the K class conditional densities" -> "the K-class conditional densities"

    "3 class naive Bayes classifier" -> "a 3-class naive Bayes classifier"

    Exercise 4.19: .. "semi tied covariances" -> "semi-tied covariances"

      "class conditional densities" -> "class-conditional densities"
          (and below)

      "the 2 class case" -> "the 2-class case"

p150 "Choosing the mode as a summary of a posterior distribution is often
      a very poor choice,"
    Style: 2x "choose". How about:
      "The mode of a posterior distribution is often a very poor choice
      as a summary,"

      "for a 1d continuous space" -> "for an 1D continuous space"

p152 "Unfortunately, optimizing Equation 5.7 is often difficult, which
      minimizes the appeal of the whole approach."
      It reduces it; I'm not sure it minimizes it!
      (Surely it could be even worse!)

p153 "This is illustrated in Figure 2.3(c)."
      Sadly, figure 2.3 as printed only has (a) and (b).

    "Fortunately, the mechanics of computing a credible interval
      is just as easy as ..."
      I would accept "mechanics" with a singular verb if it were
      the academic discipline of mechanics ("Quantum physics was
      easy, but Mechanics has always escaped me." -- as opposed
      to "mechanics have always escaped me", perhaps characterizing
      a particular revulsion I only evoke in people I ask to work
      on my car--) but with these applied mechanics,
      plural feels more natural.

p155 5.3 "Similiarly, in Figure 7.8(a), we saw ..." We're only up to
      5.5 at this point, so "saw" -> "will see" ?

p160 5.3.2.3 "Gaussian-Gaussian-Wishart model"
    Is that really what it's called, or is this one Gaussian too many?

p162 5.3.2.4 "similar expression to BIC/ MDL"
    Seriously, what's up with the spacing around slashes?

p164 "In Figure 5.9(b) shows" -> "Figure 5.9(b) shows"

    "selection/ hypothesis"
    another spurious space after a slash

p167 "Hence Jeffreys' prior"
    I applaud your use of the posessive form here -- can we apply
    it everywhere consistently? (Up to know it's been Jeffreys prior.)

p176 5.7 "We have seen how probability theory can be used to
    represent and and updates our beliefs" -> "... update ..."

p177 Equation 5.99 arg min -> argmin

p178 Equation 5.102 arg max -> argmax

p180 "object/ event" -> "object/event"

p182 Figure 5.15b What's up with the horizontal scale clipped to 0..1 in
    (a), but not in (b)? Why do these boxes have different shapes?

    The fonts labelling the axes are too small to read for me,
    and the red and blue dots the text refers to are all but invisible.

p183 "Table 5.5 gives a worked example that illustrates the difference."
    I think everybody understands you, but this is a weird
    use of "worked" to me.   Maybe just "Table 5.5 illustrates
    the difference with an example."

p185 "The intuition is that we should pick actions about which
    we believe are good, ..."
    Delete "about"?

p186 5.7.3.3 "In Setion 10.6" -> "In Section 10.6"

    "In otherwords," -> "In other words,"

    "values/ preferences" -> "values/preferences"

p187 There's a table hovering unlabeled at the top of the page that
belongs into Exercise 5.4, following "as the following loss matrix:".

Exercise 5.4 b "Now suppose P(y = 1|x) = 0.4."
The math typesetting context should last all the way to the 0.4,
not end with the closing parenthesis.

5.4 c, "decisionn" -> "decision"

5.5 "theory/ economics" -> "theory/economics"

Exercise 5.6 "the bayes factor" -> "the Bayes factor"

p189 "in the model, Compute" -> "in the model. Compute"

p194 Equation 6.6: Hm! I don't know what the subscript kk means.
Subscript k I would have expected.

p195 "theorem, that" -> "theorem that"

p197 "However, It" -> "However, it"

p200 "Since frequentist decision theory does not provide an automatic
way to choose the best estimator, we need to come up with other
heuristics for choosing amongst them."
The "them" has no direct referent in this sentence, so this is
a little iffy (even though people will understand what you mean.)
How about just s/choosing amongst them/that/ ?