620-639 DATA MINING
Errata sheet

The phrase in quotes is the slide heading closest to the slide with an error. S1 denotes the first slide under that heading, S2 the second etc. If no slide is mentioned, then it is the first slide S1 by default. PLEASE LET ME KNOW IF YOU FIND ERRORS NOT ON THIS LIST.

1. Week 3: “common loss functions” S2 - The notation not completely consistent with the rest of the course. It’s probably best to change the three expressions to:

\[ Y_i \approx f(X_i), \]
\[ E_{X_i,Y_i}[L\{Y_i, f(X_i)\}] \text{ and} \]
\[ \frac{1}{n} \sum_{i=1}^{n} L\{Y_i, f(X_i)\}. \]

2. Week 3: “proof of theorem” S1- Remove minus sign in last equation.

3. Week 3: “Parameter variances” S1- The transposition of one of the terms has gone wrong and the equation does not cancel as it should. The right hand side of the fourth line of algebra expression should have \(X(X^TX)^{-1}\) rather than \((X^TX)^{-1}X^T\), with the corresponding change in the next line too (to \((X^TX)^{-1}X^T\sigma^2X(X^TX)^{-1})\).

4. Week 4: “The inverse logistic function” S1- Should label axes \(x\) and \(g(x)\).

5. Week 4: “Fitting the logistic model” S1 - The formula for the elements of the Hessian should be negated (two changes - one to the elements plus a negative for the matrix form)

6. Week 4: “Fitting the logistic model” S2 - \(X^T\beta_t\) in the third line of algebra should be \(X\beta_t\).

7. Week 4: “Computation time” S1- This should be \(O(np^2)\) instead of \(O(n^2k)\).

8. Week 4: “Gaussian linear model” - Line three \(Y^T\) should be \(Y_i\). Also, final line should have \(-0.5\sigma^{-2}\) in place of \(-\sigma^{-2}\).

9. Week 5: “Partial dependence plots” S2 - The equation should be

\[ E\{f(X_S, X_C)|X_C = x_c\}. \]

10. Week 6: “Today’s lecture” S1 - “Generalised linear models” should be “Generalised additive models”.

11. Week 6: “Logistic basis expansion” S1 - In the last expression $\beta_m$ is written twice for no reason. $X_{ij}$ should be $X_i$.

12. Week 7: “Incorporating the response” S2 - The expression in line 6 should be

$$c_m = \frac{1}{n_m} \sum_{i; X_i \in R_m} Y_i,$$

where $n_m = |R_m|$.

13. Week 7: “Classification trees” S3 - sum should be $\sum$ when defining the Gini index.

14. Week 8: “Gradient descent” S2 - Line 10 should read

$$\theta_j^{(m)} = \theta_j^{(m-1)} - \rho \left[ \frac{\partial T(\theta)}{\partial \theta_j} \right]_{\theta = \theta^{(m-1)}}.$$

15. Week 8: “Gradient descent” S4 - “gradient boosting” should read “gradient descent” (twice).

16. Week 8: “Loss functions” S2 - Misclassification loss does not need the summation, as we just give loss for individual observations. It should thus be $\mathbb{I}[Y_i \neq \text{sign}\{f(X_i)\}]$.

17. Week 8: “More on pseudo-residuals” - The top right table element should be

$$-\frac{\partial L(Y_i, f(X_i))}{\partial f(X_i)}.$$

18. Week 8: “Shrinkage” S1 - the centred equation should not have a curly bracket on the right.

19. Week 9: “Connecting neurons” S2 - The second set of equations should have $\beta$’s instead of $\alpha$’s. Also, the first set of equations should have 3 equations, the second 4.

20. Week 9: “Connecting neurons” S5 - “The overall level of weights” appears twice in the second paragraph, and each should actually read “The overall level of parameters”.

21. Week 9: “Other comments” S1 - The $O(NpML)$ term should be $O(npML)$.

22. Week 10: “Linear decision boundaries” S1: For clarity the first expression should have brackets for the exponent:

$$\exp\{- (\alpha + \beta_1 X_{i1} + \beta_2 X_{i2})\}$$
23. Week 10: “Linear decision boundaries” S1: Last line should not have negative sign, \((\alpha + t) + \beta_1 x_1 + \beta_2 x_2 = 0\).

24. Week 10: “Linear decision boundaries” S2: In the second line, \(\alpha = -0.1\). Classification is misspelt.

25. Week 10: “Lagrangian Multipliers” S2: \(\nabla f(\beta^*, \beta_0)\) should be \(\nabla f(\beta^*, \beta_0^*)\).

26. Week 10: “Support vector classifier” S2: The expression \(Y_i(X_i^T \beta + \beta_0) \geq 1\) in the last line should have a \(-\xi_i\) as well:

\[ Y_i(X_i^T \beta + \beta_0) \geq 1 - \xi_i. \]

This modification also needs to be made where the optimisation expression appears later, in equations (7) and (8).

27. Week 10: “Support vector machines” S1: On the third last line, \(\beta_0\) should be \(\beta_0\).


29. Week 11: There was some ambiguity in the PCA section as to whether the covariance matrix \(\Sigma\) equalled \(\frac{1}{n} X^T X\) or just \(X^T X\). It was defined as the former, which corresponds to the most familiar definition of variance but not all expressions had the factor \(\frac{1}{n}\) where they should have - in particular on

- slide 36 (two times),
- point 2 of slide 38, and
- slide 47 (proof of Theorem 3).

The factor is only ever a spectator constant in the theory.