STAT 542: Homework 2

Due: Feb. 23 midnight on Canvas

Please make sure that your solutions are readable and the file size is reasonable. Typing the answers is highly encouraged.

Problem 1.

[2pts] Complete Exercise 4.2 in ESL to show the equivalence of LDA and linear regression in a certain setting. Our `Notes LDA.pdf` on Canvas already provided most of the ingredients of the calculations. Address the following question to complete the exercise:

- Let $\hat{\delta}_1(x)$ and $\hat{\delta}_2(x)$ be the discriminant functions in LDA (note that the condition in part (a) of Ex 4.2 is $\hat{\delta}_2(x) > \hat{\delta}_1(x)$). Let $\hat{f}(x)$ be the linear regression function in part (e). Show that when $N_1 = N_2$ we have

$$\hat{f}(x) = \lambda(\hat{\delta}_2(x) - \hat{\delta}_1(x)) \tag{1}$$

  for some scalar $\lambda > 0$ depending on the training data (but not on the new input x). What is λ?

Problem 2.

[2pts] Run the iris dataset example for `lda` documentation:
`https://www.rdocumentation.org/packages/MASS/versions/7.3-58.2/topics/lda`
After training the model, write codes to use the test set to compute the confusion matrix and then the values of sensitivity and specificity (see `s6_logistic.pdf` for definitions). Please include a screenshot to show your codes and results after running the code.

Hint: In the lecture we defined sensitivity and specificity for the case of binary class, but in the multiclass case as in the homework, we can use a "one-against-all" approach to extend the definition; see e.g. `https://stackoverflow.com/questions/55635406/how-to-calculate-multiclass-overall-accuracy-sensitivity-and-specificity`

Problem 3.

[3pts] Repeat Problem 2 using logistic regression instead of LDA. Compare the errors (sensitivity and specificity) with the LDA method. Also compare the performances with and without the intercept.

Problem 4.

[1 bonus point] In our slides page "Advantages of LDA over logistic regression", it was mentioned that logistic regression may be unstable in certain settings. Consider the following example: In binary logistic regression, suppose that $p > n$, X has full row rank (rank=n), and there is no intercept. Show that the likelihood function is not maximized by any finite β; in fact we can let β diverge while the likelihood tends to 1.

Hint: pick a β so that there is no training error in classification, and then set $\beta \leftarrow t\beta$ with the scalar $t \to \infty$.