Homework Assignment 4a
Due: Tuesday, Apr 2, 2024, 11:59 p.m.
Total marks: 55

Question 1. [35 MARKS]

In this question, you will implement logistic regression for binary classification. Initial code has been given to you, in `A4.jl`. You will be running on a physics data set, with 8 features and 100,000 samples (called susysubset). The features are augmented to have a column of ones (to create the bias term) in the code (not in the data file itself). We should be able to outperform random predictions, provided by a random classifier.

(a) [5 MARKS] Implement the `sigmoid` function.

(b) [10 MARKS] Implement the `get_features` function.

(c) [5 MARKS] Implement the `loss` function.

(d) [5 MARKS] Implement the `gardient` function.

(e) [5 MARKS] Implement the `update!` function.

(f) [5 MARKS] **WRITTEN:** Compare `RandomModel`, `LogisticRegressor` and `Polynomial3Model`. Report their average misclassification error. Provide 1-2 sentences commenting on the results.

Question 2. [20 MARKS]

In this question, you will use the paired t-test to compare the performance of two models. You will compare the above `LogisticRegression` model and `Polynomial3Model` model, both using RMSProp. You will run this comparison using `A4.jl`. You hypothesize that `Polynomial3Model` is better than `LogisticRegression`, and so want to run a one-tailed test to see if that is true.

(a) [5 MARKS] **WRITTEN:** Define the null hypothesis and the alternative hypothesis. Use μ_1 to be the true expected squared error for `LogisticRegression` and μ_2 the true expected squared error for `Polynomial3Model`.

(b) [10 MARKS] Now let's run the paired t-test. (Note: we should actually check for violated assumptions. We did this for you and saw that that both errors appear approximately normal distributed and have similar variance. Thus, it is acceptable to use the paired t-test). To run this test, you need to compute the p-value. Implement the `tDistPValue` function, which returns the p-value for the one-tailed paired t-test.

(c) [5 MARKS] **WRITTEN:** Report the p-value. Would you be able to reject the null hypothesis with a significance threshold of 0.05?

Homework policies:

Your assignment should be submitted as two pdf documents and a .jl notebook, on eClass. **Do not** submit a zip file with all three. One pdf is for the written work, the other pdf is generated from the .jl notebook. The first pdf should contain your answers for questions starting with "**WRITTEN:**". Your answers must be written legibly and scanned or must be typed (e.g., Latex).

This .pdf should be named Firstname_LastName_Sol.pdf, For your code, we want you to submit it both as .pdf and .jl. To generate the .pdf format of a Pluto notebook, you can easily click on the circle-triangle icon on the right top corner of the screen, called Export, and then generate the .pdf file of your notebook. The .pdf of your Pluto notebook as Firstname_LastName_Code.pdf while the .jl of your Pluto notebook as Firstname_LastName.jl. All code should be turned in when you submit your assignment.

Because assignments are more for learning, and less for evaluation, grading will be based on coarse bins. **The grading is atypical**. For grades between (1) 81-100, we round-up to 100; (2) 61-80, we round-up to 80; (3) 41-60, we round-up to 60; and (4) **0-40, we round down to 0**. The last bin is to discourage quickly throwing together some answers to get some marks. The goal for the assignments is to help you learn the material, and completing less than 50% of the assignment is ineffective for learning.

We will not accept late assignments. There is no late penalty policy. The assignments must be submitted electronically via eClass on time, by 11:59 pm Mountain time on the due date. There is a grace period of 48 hours when assignments will be accepted. No submissions will be accepted after 48 hours after the deadline, and the assignment will be considered as incomplete if not submitted.

All assignments are individual. All the sources used for the problem solution must be acknowledged, e.g. web sites, books, research papers, personal communication with people, etc. Academic honesty is taken seriously; for detailed information see the University of Alberta Code of Student Behaviour.

<div align="center">

Good luck!

</div>