COMPSCI 4ML3, Introduction to Machine Learning
Assignment 3, Winter 2021

Hassan Ashtiani, McMaster University

Due date: Sunday, March 21, 9pm

Notes. This assignment has a programming component. You should upload only two files: a pdf file for all the solutions including questions 1 (call it MacID.pdf) and a Jupyter notebook file for your codes (call it MacID.ipynb). There are 10 bonus points; if you get a grade above 100, the extra points will be used to compensate the grade of your other assignments in case they are less than 100 (the bonus points will not help with the exams).

1. **[85 points]** For this question you will use a python notebook that you can find in https://colab.research.google.com/drive/1rSHazViMkbpUMis8fyxPDe8Jvs5tM2mv?usp=sharing. In fact, this notebook is shared via Google's colab so you can just go to the File menu and "save a copy" in your own drive. Then, you can start editing the notebook in your own drive. The benefit of this approach is that you can simply run the code on the cloud and you don't need to install any packages, etc. Alternatively, you can download the notebook on your own computer and run it like assignment 2.

   Follow the instructions given in the notebook. Include the answers/graphs/analyses of the three tasks in your final pdf report. Additionally, upload your modified notebook (ipynb file) that includes your code+outcomes.

2. In the class we talked about Gaussian Discriminant Analysis for binary classification, where we considered the following probabilistic model for the data-generating distribution:

$$P(x|y=0,\mu_1,\mu_2,\beta) = \frac{1}{a}\exp -\|x-\mu_1\|_2^2$$
$$P(x|y=1,\mu_1,\mu_2,\beta) = \frac{1}{a}\exp -\|x-\mu_2\|_2^2$$
$$P(y=0|\mu_1,\mu_2,\beta) = 1-\beta$$

   (a) **[10 points]** Given a sample $Z = \{(x^i, y^i)\}_{i=1}^N$, what is the maximum likelihood estimate for μ_1, μ_2, and β? You don't need to redo the whole proof; you can reuse the parts that we have shown in the class.

   (b) **[15 points]** Assume that we are given μ_1, μ_2, and β, and want to use our model for classification. What will be the shape of the decision boundary? Prove your result.