Assignment 2: Classification

Due October 21 at 11:59pm
60 marks total

This assignment is to be done individually.

Important Note: The university policy on academic dishonesty (cheating) will be taken very seriously in this course. You may not provide or use any solution, in whole or in part, to or by another student.

You are encouraged to discuss the concepts involved in the questions with other students. If you are in doubt as to what constitutes acceptable discussion, please ask! Further, please take advantage of office hours offered by the instructor and the TA if you are having difficulties with this assignment.

DO NOT:

- Give/receive code or proofs to/from other students
- Use Google to find solutions for assignment

DO:

- Meet with other students to discuss assignment (it is best not to take any notes during such meetings, and to re-work assignment on your own)
- Use online resources (e.g. Wikipedia) to understand the concepts needed to solve the assignment

---

1 Softmax for Multi-Class Classification (10 marks)

The *softmax function* is a multi-class generalization of the logistic sigmoid:

$$p(\mathcal{C}_k|\boldsymbol{x}) = \frac{\exp(a_k)}{\sum_j \exp(a_j)} \tag{1}$$

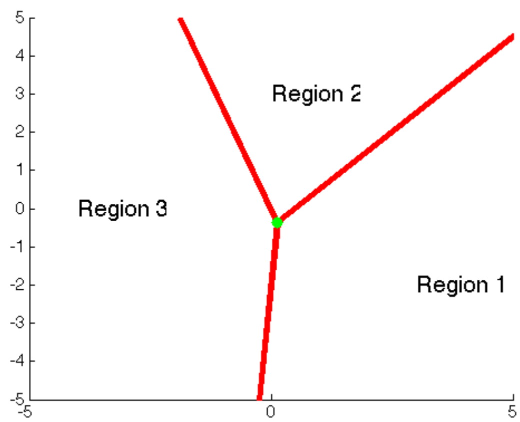
Consider a case where the *activation functions* a_j are linear functions of the input. Assume there are 3 classes ($\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$), and the input is $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$

$$a_1 = 3x_1 + 1x_2 + 1$$

$$a_2 = 1x_1 + 3x_2 + 2$$

$$a_3 = -3x_1 + 1.5x_2 + 2$$

The image below shows the 3 decision regions induced by these activation functions, their common point intersection point (in green) and decision boundaries (in red).



Answer the following questions. For 2 and 3, you may provide qualitative answers (i.e. no need to analyze limits).

1. (3 marks) What are the probabilities $p(\mathcal{C}_k|\boldsymbol{x})$ at the green point?
2. (3 marks) What happens to the probabilities along each of the red lines? What happens as we move along a red line (away from the green point)?
3. (4 marks) What happens to the probabilities as we move far away from the intersection point, staying in the middle of one region?

2 Generalized Linear Models for Classification (10 marks)

Consider a generalized linear model for classification over a two-dimensional input $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$. Assume there are 2 classes.

Use the polynomial basis functions below:

$$\phi_0(\boldsymbol{x}) = 1$$

$$\phi_1(\boldsymbol{x}) = x_1^2$$

$$\phi_2(\boldsymbol{x}) = x_2^2$$

Suppose we learn weight vectors $\boldsymbol{w}^1 = (6, 2, 2)$ and $\boldsymbol{w}^2 = (8, 0, 0)$, so that

$$y_1(\boldsymbol{x}) = 6\phi_0(\boldsymbol{x}) + 2\phi_1(\boldsymbol{x}) + 2\phi_2(\boldsymbol{x})$$

$$y_2(\boldsymbol{x}) = 8\phi_0(\boldsymbol{x}) + 0\phi_1(\boldsymbol{x}) + 0\phi_2(\boldsymbol{x})$$

Answer the following questions.

1. (3 marks) Draw the function $y_1(\boldsymbol{x})$.
2. (3 marks) Draw the function $y_2(\boldsymbol{x})$.
3. (4 marks) Draw the decision regions that are given by the multi-class method that assigns a label to a point $\boldsymbol{x}$ by $\arg\max_k y_k(\boldsymbol{x})$.

3 Logistic Regression (40 marks)

In this question you will examine optimization for logistic regression.

1. Download the assignment 2 code and data from the website. Run the script `logistic_regression.py` in the lr directory. This code performs gradient descent to find $\boldsymbol{w}$ which minimizes negative log-likelihood (i.e. maximizes likelihood).

   Include the final output of Figures 2 and 3 (plot of separator path in slope-intercept space; plot of neg. log likelihood over iterations) in your report.

   Why are these plots oscillating? Briefly explain why in your report.

2. Create a Python script `logistic_regression_mod.py` for the following.

   Modify `logistic_regression.py` to run gradient descent with the learning rates $\eta = 0.5, 0.3, 0.1, 0.05, 0.01$.

   Include in your report a single plot comparing negative log-likelihood versus iteration for these different learning rates.

   Compare these results. What are the relative advantages of the different rates?

3. Create a Python script `logistic_regression_sgd.py` for the following.

   Modify this code to do stochastic gradient descent. Use the parameters
   $\eta = 0.5, 0.3, 0.1, 0.05, 0.01$.

   Include in your report a new plot comparing negative log-likelihood versus iteration using stochastic gradient descent.

   Is stochastic gradient descent faster than gradient descent? Explain using your plots.

4. Create a Python script `logistic_regression_irls.py` for the following.

   Modify this code to use iterative reweighted least squares (IRLS, Eqn. 4.99). The Python function `np.diag` is useful for Eqn. 4.98.

   Note that this only takes about 3 lines of code to implement. If you're doing more work, stop, read the textbook, or ask me or the TAs for help.

   Include new plots of Figures 2 and 3 using IRLS in your report.

   Yes, it is that fast.

Submitting Your Assignment

The assignment must be submitted online at `https://courses.cs.sfu.ca`. You must submit three files:

1. An assignment report in **PDF format**, called `report.pdf`. This report must contain the solutions to questions 1-2 as well as the figures / explanations requested for 3.
2. A .zip file of all your code, called `code.zip`.