Assignment 2: Classification

Due October 20 at 11:59pm
89 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

---

Question 1 (9 marks)

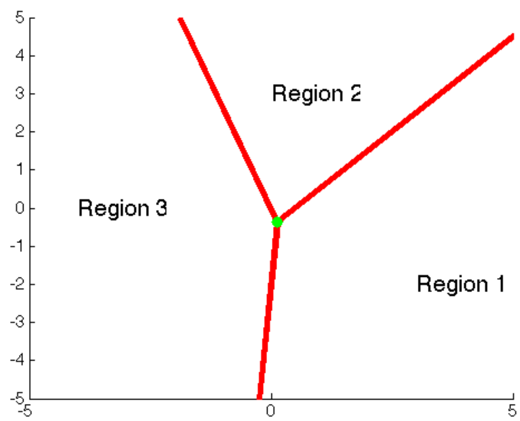
The *softmax function* is a multi-class generalization of the logistic sigmoid:

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

Consider a case where the *activation functions* a_j are linear functions of the input:

- 3 classes $(\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3)$
- $a_1 = 3x_1 + 1x_2 + 1$
- $a_2 = 1x_1 + 3x_2 + 2$
- $a_3 = -3x_1 + 1.5x_2 + 2$

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. (3 marks) What happens to the probabilities as we move far away from the intersection point, staying in the middle of one region?

Question 2 (10 marks)

Show that the exponential kernel $k(\boldsymbol{x}, \boldsymbol{x}') = \exp(k_1(\boldsymbol{x}, \boldsymbol{x}'))$ (Eqn. 6.16) corresponds to a dot product in an infinite dimensional feature space.

- Start by writing down exp as a power series (Taylor expansion around 0).
- You may then make use of Eqn. 6.15, which states that for a polynomial $q(k_1(\boldsymbol{x}, \boldsymbol{x}'))$, e.g. $a_d k_1(\boldsymbol{x}, \boldsymbol{x}')^d$ with $a_d > 0$, there exists a feature space $\boldsymbol{\phi}_d(\boldsymbol{x})$ such that $q(k_1(\boldsymbol{x}, \boldsymbol{x}'))$ acts as a dot product in that space.
- Write down the infinite dimensional space in which $\exp(k_1(\boldsymbol{x}, \boldsymbol{x}'))$ corresponds to a dot product (using the spaces from above).

Question 3 (15 marks)

In some instances, we may wish to use soft-margin support vector machines with different slack tradeoff parameters C_1 and C_2 for positive and negative training examples respectively.

For example, in imbalanced datasets it is common to reweight the positive and negative examples according to their frequencies; or to express different "penalty" for mis-classifying positives versus negatives. The result is the following optimization problem:

$$\min_{\boldsymbol{w},b} \frac{1}{2}||\boldsymbol{w}||^2 + C_1 \sum_{n \in \mathcal{P}} \xi_n + C_2 \sum_{n \in \mathcal{N}} \xi_n \tag{2}$$

$$\text{s.t. } t_n y(\boldsymbol{x}_n) \geq 1 - \xi_n \ \forall n \tag{3}$$

$$\xi_n \geq 0 \ \forall n \tag{4}$$

where $\mathcal{P}$ and $\mathcal{N}$ contain the indices of the positive and negative training examples respectively.

Derive the Langrangian dual form ($\tilde{L}(\boldsymbol{a})$) and *box constraints* for this problem.

Question 4 (25 marks)

In this question you will compare 3 methods for optimization for logistic regression.

1. Download the assignment 2 code and data from the website. Run the script `logistic_regression.m` in the lr directory.

   This code performs gradient descent to find $\boldsymbol{w}$ which maximizes the likelihood (more precisely, minimizes negative log-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.**

   How might you fix this? **Fix this, and include new plots in your report and an explanation of your fix.**

2. Create a MATLAB script `logistic_regression_sg.m` for the following.

   Modify this code to do stochastic gradient descent. Use the same parameter $\eta = 0.003$

   Include new plots of Figures 2 and 3 using stochastic gradient descent in your report.

3. Create a MATLAB script `logistic_regression_irls.m` for the following.

   Modify this code to use iterative reweighted least squares (IRLS, Eqn. 4.99). The built-in MATLAB function `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 TA for help.

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

   Yes, it is that fast.

Question 5 (30 marks)

In this question you will use support vector machines for image classification[1]. The data are in the tarball on the website (image directory).

The training data are in `train.mat`. The test data are in `test.mat`. A feature vector for each image has been computed for you[2].

Each .mat file contains:

- Ftrain (or Ftest): an n_{images}-by-M matrix. Each row is the M-dimensional representation for an image.
- Ltrain (absent for test): an n_{images}-by-1 vector. Each entry is the class label (in $\{1, 2, \ldots, 10\}$) – a 10-class problem.
- C: a 10-by-1 cell array with class names (e.g. `C{1}` is abbey)
- Itrain (or Itest): an n_{images}-by-1 cell array of relative image locations[3].

Download libsvm, which has a MATLAB interface on it, from: `http://www.csie.ntu.edu.tw/~cjlin/libsvm/#matlab`

Instructions for installation are available on that website.

Experiment with different kernels and values for parameter C, using cross-validation on the training data in `train.mat`.

Choose what you think is the best classifier, then run it on the unlabeled data in `test.mat`. Produce an output vector `Ptest` that is 500-by-1 of class labels (i.e. a number in $\{1, 2, \ldots, 10\}$).

Save the vector `Ptest` and your login to identify you in a file imagetest.mat:

```
Ptest = ...
name = 'your login';
save('imagetest.mat','Ptest','name');
```

There is also a MATLAB function to create a webpage to visualize your classification results:

```
webpageDisplay(Itest,Ptest,C);
or
webpageDisplay(Itrain,Ptrain,C);
```

Open the webpage `output.html` this creates to see your results. Note that you shouldn't tune your algorithm based on the *test* results.

[1]The data come from the SUN Database `http://vision.cs.princeton.edu/projects/2010/SUN/`.
[2]These are *dense SIFT* features.
[3]The base URL is `http://labelme.csail.mit.edu/Images/users/antonio/static_sun_database`.

Include a few plots showing cross-validation results, and describe the kernels with which you experimented, in your report. State which kernel/parameter values you used for producing `Ptest`.

Bonus marks and a prize will be given to the student(s) with the best classification performance!

Submitting Your Assignment

The assignment must be submitted online at `https://courses.cs.sfu.ca`. In order to simplify grading, you must adhere to the following structure.

Submit:

1. An assignment report in **PDF format**, called `report.pdf`. This report must contain the solutions to questions 1-3 as well as the figures / explanations requested for 4-5.

2. You must upload the following files:

   (a) `logistic_regression_sg.m`

   (b) `logistic_regression_irls.m`

   (c) `imagetest.mat`