Problem Set 2

Submission instructions and due date:

The due date for this homework assignment is as listed on Blackboard as part of the "Problem Set ..." file name on Blackboard. The completed assignment must be submitted in class (on paper and stapled according to the instructions below) at the beginning of the class session on the above due date, unless a different due date is announced in class or on Blackboard.

What to submit (print and staple in this order):

1) hw2scores.txt (complete and print out).
2) solutions to written problems.
3) solutions to programming problem.

For item 2 of the list above (solutions to written problems) – solutions to written problems can be handwritten; however, they should be written out clearly. It is required that you show all relevant steps leading to the solutions.

For item 3 of the list above (programming problem) – submit printouts of the completed code, any screenshots of the output related to this assignment and the results' plots.

Grading:

This problem set is self-graded. It is strongly recommended that you solve and complete all of its tasks (written and programming), since – in addition to contributing towards the homework component of the final course-grade – they should help you prepare for quizzes/exams. You must also submit your self-assigned grades in "hw2scores.txt" (see also instructions within that file). To verify self-grading, for each homework assignment we will randomly choose a subset of submissions for grade checking and re-grade them if needed. Reminders: Late-homework policy is as listed in the course syllabus. Copying solutions from any source – e.g., online, solutions manuals, assignment solutions from prior semesters, or solutions from any other sources – is prohibited.

1 Written Problems

1.1 Logistic sigmoid function [20 points].

Show that the logistic sigmoid function $\sigma(a) = \frac{1}{1+exp(-a)}$ satisfies the property $\sigma(-a) = 1 - \sigma(a)$ and that its inverse is given by $\sigma^{-1}(y) = \ln\{y/(1-y)\}$.

1.2 Derivation of LDA [20 points].

Show that the log-odds decision function a(x) for LDA

$$a = \ln \frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)}$$

is linear in x, that is, we can express $a(x) = \theta^T x$ for some θ. Show all your steps.

1.3 LR Classification with Label Noise [20 points].

Suppose you are building a logistic regression classifier for images of dogs, represented by a feature vector x, into one of two categories $y \in \{0, 1\}$, where 0 is "pug" and 1 is"bulldog". You decide to use the logistic regression model $p(y = 1|x) = h_\theta(x) = \theta^T x$. You collected an image dataset $\mathbf{D} = \left\{x^{(i)}, t^{(i)}\right\}$, however, you made some mistakes in assigning labels $t^{(i)}$. You estimate that you were correct in about τ fraction of all cases.

(a) Write down the equation for the posterior probability $p(t = 1|x)$ of the label being 1 for some point x, in terms of the probability of the true class, $p(y = 1|x)$.

(b) Derive the modified cost function in terms of θ, $x^{(i)}$, $t^{(i)}$ and τ.

2 Programming Problem: Regularized Logistic Regression [40 points].

Complete the code in the Regularized_regression.py file. Additional information is provided within the Regularized_regression.py file. For this particular exercise, your code additions must implement the required computations explicitly, i.e., *without* calls to machine-learning toolkits that provide the functionality you are asked to add within the Regularized_regression.py file. Once you complete the code, you will plot the decision boundary. Your final result should be similar to: