HMC CS 158, Fall 2017
Problem Set 4 Exercises: Logistic Regression, Perceptron

Goals:
- To prove that gradient descent is an appropriate optimization technique for logistic regression.
- To investigate perceptron feasibility for different data sets.

Submission

You should submit any answers to the exercises in a single file `writeup.pdf`. This writeup should include your name and the assignment number at the top of the first page, and it should clearly label all problems. Additionally, cite any collaborators and sources of help you received (excluding course staff), and if you are using slip days, please also indicate this at the top of your document.

1 Logistic Regression [8 pts]

Consider the cost function for logistic regression, which minimizes the negative log probability of the data:

$$J(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \left[y^{(i)} \log h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right)\right)\right]$$

In this exercise, you will show that $J(\boldsymbol{\theta})$ has no local minima other than the global one. Thus, gradient descent (with an appropriately chosen step size) is guaranteed to converge to the global minimum, making it an appropriate optimization technique for logistic regression.

(a) **(2 pts)** Start by finding the partial derivatives $\frac{\partial J}{\partial \theta_j}$.

(b) **(2 pts)** Next find the partial second derivatives $\frac{\partial^2 J}{\partial \theta_j \partial \theta_k}$.

(c) **(4 pts)** Finally, show that J is a convex function and therefore has no local minima other than the global one.
Hint: A function J is convex if its Hessian (the matrix $\boldsymbol{H}$ of second derivatives with elements $H_{jk} = \frac{\partial^2 J}{\partial \theta_j \partial \theta_k}$) is positive semi-definite (PSD), written $\boldsymbol{H} \geq 0$. A matrix is PSD if and only if

$$\boldsymbol{z}^T \boldsymbol{H} \boldsymbol{z} \equiv \sum_{j,k} z_j z_k H_{jk} \geq 0.$$

for all real vectors $\boldsymbol{z}$.

Parts of this assignment are adapted from course material by Andrew Ng (Stanford), Jenna Wiens (UMich), and Tommi Jaakola (MIT).

2 Perceptron [2 pts]

Design (specify $\boldsymbol{\theta}$ for) a two-input perceptron (with offset) that computes the following boolean functions. Assume $T = 1$ and $F = -1$. If a valid perceptron exists, show that it is not unique by designing another valid perceptron (with a different hyperplane, not simply through normalization). If no perceptron exists, state why.

(a) **(1 pts)** AND

(b) **(1 pts)** XOR