Notes: (1) These questions require thought, but do not require long answers. Please be as concise as possible. (2) For problems that require programming, please include in your submission a copy of your code (with comments) and any figures that you are asked to plot and submit your code and figures electronically to BLS before the due date.

Logistic regression: Consider the objective function in logistic regression problem:

$$l(\theta) = \sum_{i=1}^{m} (y_i \log \sigma(\theta^T x_i) + (1 - y_i) \log(1 - \sigma(\theta^T x_i))),$$

where $\sigma(z) = (1 + e^{-z})^{-1}$ is the logistic function. Derive the gradient and Hessian matrix of $l(\theta)$.

(a) On the BLS system, the files `q1x.dat` and `q1y.dat` contain the inputs ($x_i \in R^2$) and outputs ($y_i \in \{0, 1\}$) respectively for a binary classification problem, with one training example per row.

(b) Implement the gradient descent method for optimizing $l(\theta)$, and apply it to fit a logistic regression model to the data. Initialize gradient descent method with $\theta = 0$ (the vector of all zeros). What are the coefficients θ resulting from your fit? (Remember to include the intercept term.)

(c) Implement Newton's method to maximize $l(\theta)$, and compare the overall running time and number of iterations needed to converge to the same precision.

(d) Plot the training data (your axes should correspond to the two coordinates of the inputs, and you should use a different symbol for each point plotted to indicate whether that example had label 1 or 0). Also plot on the same figure the decision boundary fit by logistic regression. (i.e., this should be a straight line showing the boundary separating the region where $\sigma(\theta^T x) > 0.5$ from where $\sigma(\theta^T x) < 0.5$.)