CECS 551: Adv Artificial Intelligence (Spring 2019)

Homework #1
Due 2/27/2019

---

1. You need to submit a report in hard-copy before lecture and your code to BeachBoard.
2. Hard-copy is due in class before lecture and electronic copy is due 11:59PM on BeachBoard on the due date.
3. Unlimited number of submissions are allowed on BeachBoard and the latest one will be graded.

---

1. (25 points) In logistic regression, the objective function can be written as

$$E(w) = \frac{1}{N}\sum_{n=1}^{N} \ln\left(1 + e^{-y_n w^T x_n}\right).$$

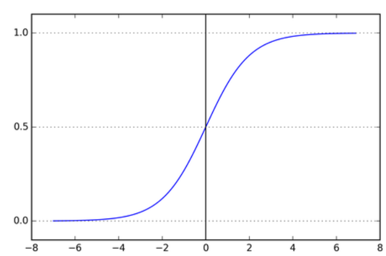
Please

(a) (10 points) Compute the first-order derivative $\nabla E(w)$. You will need to provide the intermediate steps of derivation.

(b) (10 points) Once the optimal w is obtain, it will be used to make predictions as follows:

$$\text{Predicted class of } x = \begin{cases} 1 & \text{if } \theta(w^T x) \geq 0.5 \\ -1 & \text{if } \theta(w^T x) < 0.5 \end{cases}$$

where the function $\theta(z) = \frac{1}{1+e^{-z}}$ looks like



Explain why the decision boundary of logistic regression is still linear, though the linear signal $w^T x$ is passed through a nonlinear function θ to compute the outcome of prediction.

(c) (5 points) Is the decision boundary still linear if the prediction rule is changed to the following? Justify briefly.

$$\text{Predicted class of } x = \begin{cases} 1 & \text{if } \theta(w^T x) \geq 0.9 \\ -1 & \text{if } \theta(w^T x) < 0.9 \end{cases}$$

(d) (10 points) In light of your answers to the above two questions, what is the essential property of logistic regression that results in the linear decision boundary?

2. (25 points) In softmax regression (Multinomial Logistic Regression), we compute the probability $P(y = k|x)$ for each value of $k = 1, 2, ..., K$ as

$$\begin{bmatrix} P(y=1|x) \\ P(y=2|x) \\ . \\ . \\ . \\ P(y=K|x) \end{bmatrix} = \frac{1}{\sum_{j=1}^{K} exp(\theta_j^T x)} \begin{bmatrix} exp(\theta_1^T x) \\ exp(\theta_2^T x) \\ . \\ . \\ . \\ exp(\theta_K^T x) \end{bmatrix},$$

   where θ_j are the parameters.
   Show that when $K = 2$, softmax regression reduces to the two-class logistic regression problem.

3. (50 points) **Logistic Regression for Handwritten Digits Recognition**: Implement logistic regression for classification using gradient descent to find the best separator. The handwritten digits files are in the "data" folder: train.txt and test.txt. The starting code is in the "code" folder. In the data file, each row is a data example. The first entry is the digit label ("1" or "5"), and the next 256 are grayscale values between -1 and 1. The 256 pixels correspond to a 16×16 image. You are expected to complete your solution based on the given codes. You need to complete is the "solution.py" and "main.py" files. The 'main.py' is used to test your solution. Note that code is provided to compute a two-dimensional feature (symmetry and average intensity) from each digit image; that is, each digit image is represented by a two-dimensional vector before being augmented with a "1" to form a three-dimensional vector as discussed in class. These features along with the corresponding labels should serve as inputs to your logistic regression algorithm.

   Deliverable: You should submit the "solution.py" and "main.py" files to the BeachBoard.

   Note: Please read the "Readme.txt" file carefully before you start this assignment.