CS-5331 (4331) Assignment 3

Due on 04/18/2025, submit to Blackboard

Review the demo on adversarial examples on logistic regression and complete this assignment. **Please submit a PDF version of your completed jupyter notebook.**

In the demo, each data record is labeled as either 0 or 1, and the loss function is defined as cross entropy. Recall that if we label each data sample using "-1" or "1", the loss function can be shown as below (which you just proved in Assignment 2)

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{M} \log(1 + \exp(-y_i \theta^T x_i)).$$

Consider this new loss function and implement the following in the provided jupyter notebook.

1. **(25 points)** Implement the new loss function and evaluate the empirical loss (the first highlighted cell in the provided PDF notebook).

   [Hint: use $y = 2y - 1$ to convert 0/1 label into -1/+1 lablel.]

2. **(25 points)** Implement $\frac{\partial J(\theta)}{\partial \theta}$, i.e., the derivative of the new loss function w.r.t. the model, and evaluate the result (the second highlighted cell in the provided PDF notebook).

3. **(50 points)** Implement the Fast Gradient Sign Method (FGSM) to generate adversarial examples, and demonstrate that the perturbation causes training examples to cross the decision boundary of a logistic regression model, resulting in incorrect predictions (the third and fourth highlighted cell in the provided PDF notebook).