HINT FILE: Assignment 3
(Logistic Regression and Regularisation)

Artificial Intelligence (CSE-241N)
IIT (BHU) Varanasi

27th January 2020

1 Logistic Regression

In this document we'll describe the equations for implementing a Logistic Regression model.

The predict function for logistic regression looks like:

$$y_{[n\times 1]} = sigmoid(x_{[n\times f]}W_{[f\times 1]}) \tag{1}$$

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{2}$$

where y is the predicted values for the input data x and W is the weight vector. The dimensions of the elements are in terms of n: the number of data points and f: length of feature for each data point. We'll use $\hat{y}_{[n\times 1]}$ to represent the actual values corresponding to the data points x.

The *sigmoid* function is applied element-wise.

Now we'll give the loss equation:

$$L = \frac{1}{n}\sum_{i=1}^{n} -\hat{y}_i log(y_i) - (1-\hat{y}_i)log(1-y_i) \tag{3}$$

The gradient descent update equation is given by:

$$W_{i+1} = W_i - \eta\frac{\partial L}{\partial W} \tag{4}$$

where η is the learning rate, and

$$\frac{\partial L}{\partial W} = \frac{1}{n}x^\top(y-\hat{y}) \tag{5}$$

therefore,

$$W_{i+1} = W_i - \eta\frac{1}{n}x^\top(y-\hat{y}) \tag{6}$$

2 Regularisation

Following is the formulation for L2 regularisation for any generic loss function L,

$$\mathcal{L}(W,x) = L(W,x) + \frac{\lambda}{2f}\sum_{i=1}^{f} W_i^2 \tag{7}$$

Where λ is the regularisation parameter. The gradients of weights will change in the following way

$$\frac{\partial \mathcal{L}}{\partial W} = \frac{\partial L}{\partial W} + \frac{\lambda}{f} W \tag{8}$$

Therefore, the final update becomes,

$$W_{i+1} = W_i - \eta(\frac{1}{n} x^\top (y - \hat{y}) + \frac{\lambda}{f} W_i) \tag{9}$$