**Karan Agarwalla – 180050045**

**Problem 1.1(b)**

There are four features for counts of pixels with 1, 2, 3 and 4 black neighbors. The 5th feature is the bias term (1). The features are normalized except for the fifth feature. In the implementation the fifth feature is the first and the other four follows. Accuracy observed is 100%.

The features designed must be independent of orientation and translation of figure. Hence features dependent on means and standard deviation of x-coordinate and y-coordinate that I tried didn’t give high accuracy. The normalized counts are independent of the orientation hence, give high accuracy. The combination of these features can be used to represent other metrics like area (nearly the summation of the four features).

**Problem 2**

Problem 2.1

(a)

Hence sigmoid probability in case of binary logistic regression is a special case of multi-class regression with weight vector w = w1 – w­0. Also y(i)in case of logistic regression is simply y1(i) . Hence, simplifying the expression,

which is the cross-entropy loss function used to train binary logistic regression.

(b)

(follows from (1))

Let’s define , i.e., probability of kth label on ith example

Hence,

where Pj is vector of size N\*1 whose entries are Pij­ as its entries for each example i. Similarly, Yj is vector of size N\*1 whose entries are as its entries for each example

Hence,

where P and Y are matrices of size N\*K where K is number of classes and is matrix of size N\*F where F is number of features.