Homework #6

Problem 1 (25 points)

Using the breast cancer dataset, let's encode the `diagnosis = M` as 1, and `diagnosis = B` as 0, and treat this as a binary outcome (*i.e.*, dependent variable). Let's use 3 features, namely, `smoothness_mean`, `texture_mean`, `perimeter_mean`, to build a neural network (NN). For this NN, let's have one hidden layer, together with the final layer (output layer). For the hidden layer, let's have 5 neurons, and use ReLU as the activation function. For the final layer, let's use the sigmoid function as the activations.

1. [5 points] How many parameters are there in this NN?
2. [15 points] Using the matrix notation convention shown in the class (please refer to lecture notes from week 9b, noticeably slide 3), if we have both the matrices $W^{[1]}$ and $W^{[2]}$ filled with the value of 1, and $B^{[1]}$ and $B^{[2]}$ filled with the value of 0.1. After one forward propagation, what is the value of the average prediction?

- Note that if there are `n` samples, the **average prediction** is taken over all `n` samples.
- Here we will use all samples, namely, no need for the "train/test" split.
- Please standardize the input (*i.e.*, for each feature, subtract its mean, and divide by its standard deviation) before the forward propagation.

3. [5 points] Following part 2 above, if we use logloss as the error metric, what is the value of this error metric after one forward propagation? Note here logloss is also taken as the average over all `n` samples.

> Hint: you might find this notebook helpful.

Problem 2 (10 bouns points)

Prove that, in Lloyd's algorithm for K-means clustering, for a given cluster C with n observations, the averaged pairwise within-cluster squared L2 norm equals to the sum of squared L2 norm of each point (in the cluster) to its cluster center. That is:

$$\frac{1}{n}\sum_{i,j\in C}||x_i - x_j||_2^2 = 2\sum_{i\in C}||x_i - \mu||_2^2,$$

where $\mu = \frac{1}{n}\Sigma_{i\in C}x_i$, is the centroid of the cluster.