Project Report 5

Question1:
1. Combine the training data and corresponding label together for future shuffling

2. Iterate a certain number of epochs(num_epochs)

3. In each epoch, shuffle the training data and corresponding labels. Divide these data into batches(based on given batch_size)

4. For each batch of training data, calculate the gradient of the weights(which is a column vector)

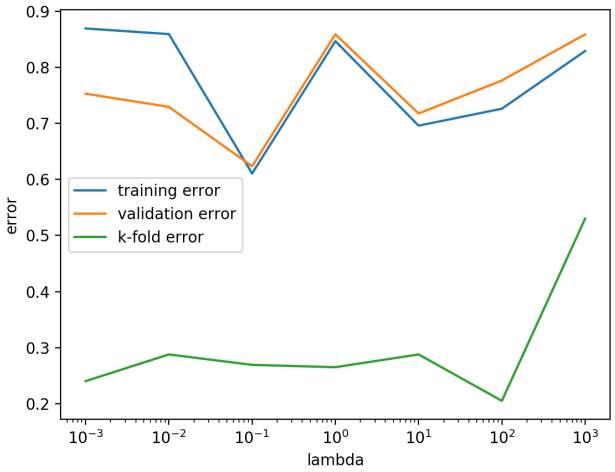
5. To calculate the gradient of weights, first calculate $h(< w, x >)$ by calling sigmoid_function. Then use the formation $\frac{1}{m}\sum_{i=1}^{m}(h(x_i) - y_i)x_{ij} + 2 * \lambda * w_j$ to calculate the gradient of w_j. The λ here is the parameter for regulation.

6. Update the weights by gradient descent algorithm

7. After the all epochs are finished, the model is ready

Question2:
If the Tikhonov regularization was used in Homework3, the model could be less overfitting since the hypothesis selected could be more simpler. The accuracy on test data could be higher.

Question3:
Here is the figure produced by function plotError().



I chose 1 as my optimal lambda. Because the accuracy is relatively high compared with other values of lambda. The k-fold error is relatively small. Also, the value of 1 is neither too large nor too small to include the regulation effect.

Question4:
No, I don't think we should account for multiple samples from one single patient. The features of the dataset are mainly computed from a digitized image. Therefore, we can assume different samples are i.i.d, because there is no strong relation between each sample from the same patient.