Machine Learning - Homework 3
Logistic Regression with Regularization
Rubiya Tasnim and Yaffa Atkins

Report:

Logistic regression with regularization:

Algorithm:

For iteration t= 0,1,2,3…………
1) Initialize W = 0
2) Calculate weight score using training data: $s = \sum_{i=0}^{d} w_i x_i$
3) First we computed sigmoid using training data and here x= s, weightscore :

$$S(x) = \frac{1}{1+e^{-x}}$$

4) We then computed gradient descent:

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)}$$

5) Then we update w = Winitial - (learning rate)*gradient descent.
6) Repeat 2-4 till enough iterations to get W
7) Loss function

L2: $\frac{\lambda}{2} \|\mathbf{w}\|^2 = \frac{\lambda}{2} \sum_{j=1}^{m} w_j^2$

$$J(\mathbf{w}) = \sum_{i=1}^{n} \left[-y^{(i)} \log\left(\phi(z^{(i)})\right) - \left(1 - y^{(i)}\right) \log\left(1 - \phi(z^{(i)})\right) \right] + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

8) Fit:
   When fitting the validation data, we minimize the weights by using gradient descent to find the quickest path toward zero. When the weights are minimized this consequently minimizes the loss because loss references weight.

9) Predict:
   Next we predicted the outputs of the x validation data using the ideal weight. We compared our predicted y with the y validation data to check our model accuracy.

5-fold Cross Validation:

Training and Testing dataset:

We used build-in function K.split() for splitting the entire D into 5 different Ds for Validation and Training.

Feature Transformation:

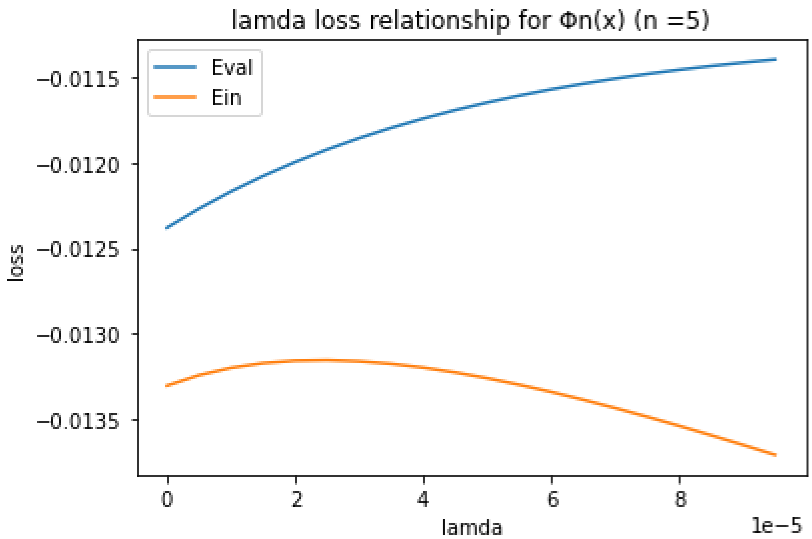
We transformed the feature space to higher dimensions with various $\Phi n(x)$ (n = 5 and n=6)
We used build-in polynomial feature transform for that.

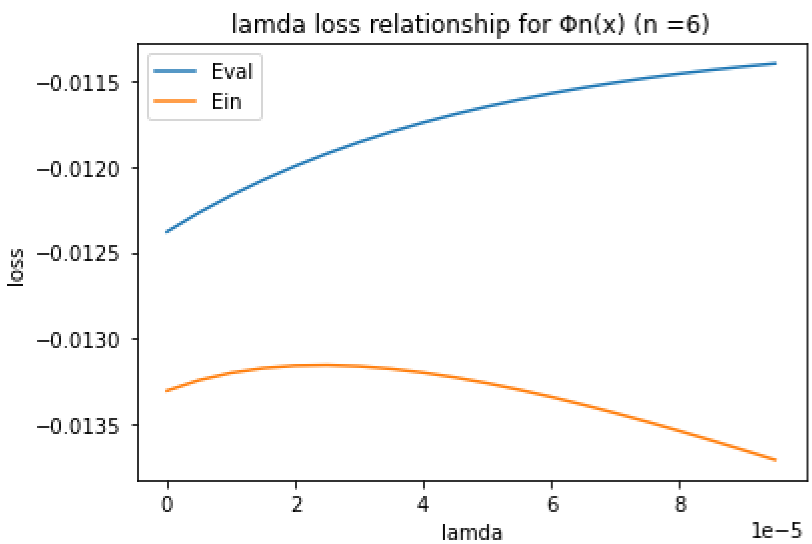
We normalized the data using this formula: (x-x.mean)/x.std

Experiment:

We picked different λ values and using these λ values we calculated the loss function using validation data to find Eval. We also calculated Ein using train data . We also used different degrees of polynomial transformations (n= 5, 6). We plot the results as shown in the plots below:

Result:





Observations:
For Eval, the small lamda is the best choice, but for Ein around 3e-5 is the choice for lamda.
Eval goes as lamda goes up.

For n=5 and n= 6 feature polynomial transformation, the plots look the same.
n>7, computer do not have enough memory to compute.

Discussion: We observed that for Eval, loss is minimized as lamda is decreased. According to our results, without regularizations we achieve the best Eval. But we are aware that regularization is for minimizing the issues with overfitting. Also, we observed our Eval returns negative values in the plot. We think our model is not accurate since we were expecting Eval decreases when regularization is applied to tackle the issues of overfitting.

For feature transformation, slightly higher n takes significantly longer to run and we ran into storage errors. The graphs do not look very different for 5 vs 6 feature transformations and that could be because the specific features added to not impact y.