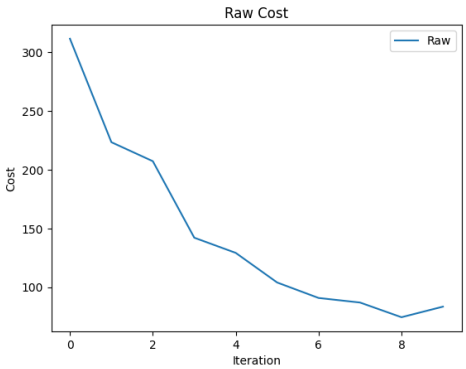
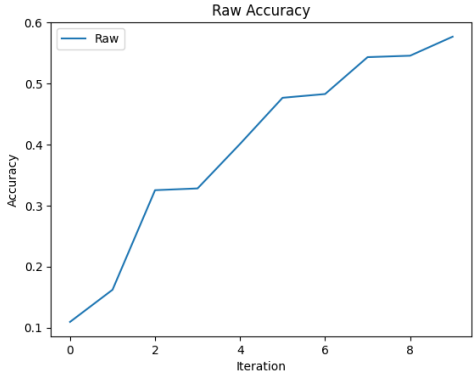
# COMPSCI 589 Machine Learning
# Assignment 7 Report

## Task 1

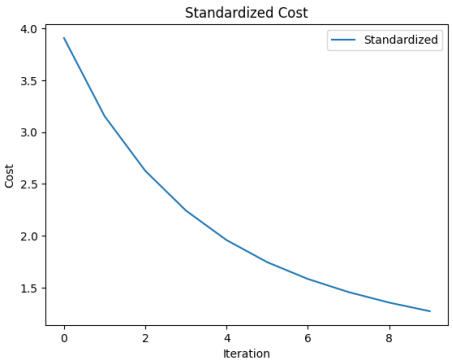
- You should return 5 plots: 1 for each cost individually, 1 with all the costs together, 1 with all the accuracies.
    - Plot of the cost functions over the 10 runs
    - Plot of the accuracies over the 10 runs

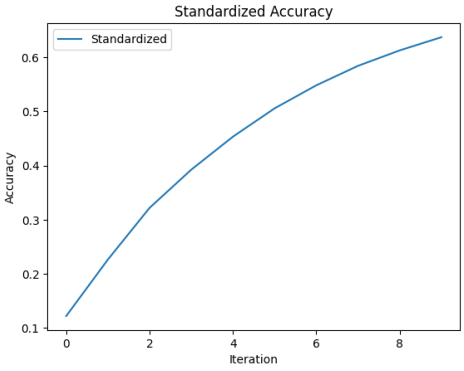
1. Original Cost and Accuracy Graphs



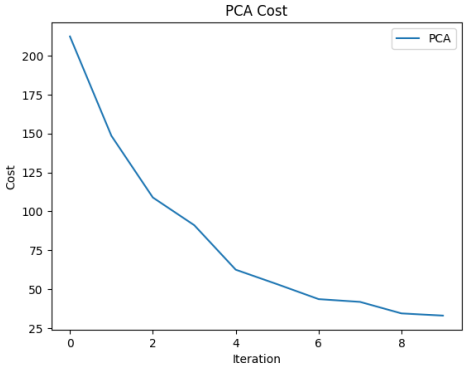



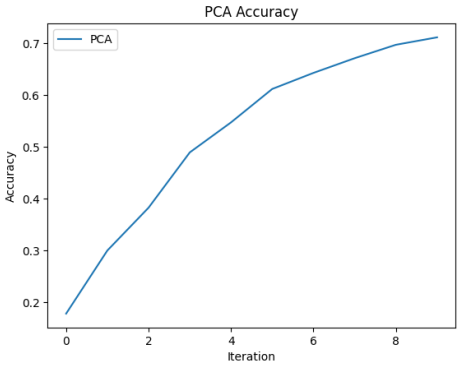

2. Standardized Cost and Accuracy Graphs



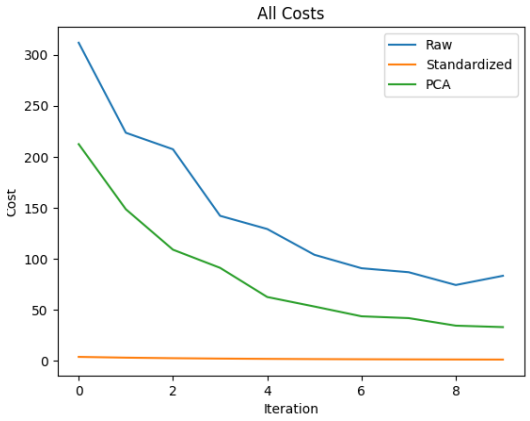




3. PCA Cost and Accuracy Graphs

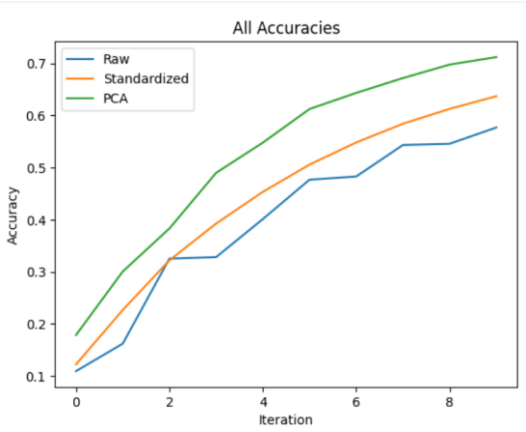


4. All costs Graph




5. All Accuracies graph




- Choice of gamma for:
  - Raw data without normalization **= -3**
  - Standard normalized data **= -1**
  - PCA-sphered data **= -2**

The gamma parameter is selected by iterating through a range of gamma values, applying gradient descent for each value, and choosing the gamma that yields the minimum cost for each respective method.

- An explanation of what you see

- Upon analysing the accuracy histories, it is evident that PCA outperforms standard normalization. The higher accuracy of the PCA-normalized data compared to standard normalization suggests that the PCA transformation captures and retains more relevant information for the classification task. PCA might be effective in extracting discriminative features and reducing dimensionality.
- Both data transformations exhibit higher accuracies compared to using the original data as is.
- In terms of cost histories, it is notable that the original data achieves a lower minimum cost compared to the other methods.
- The fact that different normalization techniques require different optimal step lengths (gamma) highlights the sensitivity of gradient descent to the choice of this hyperparameter.

# Task 2

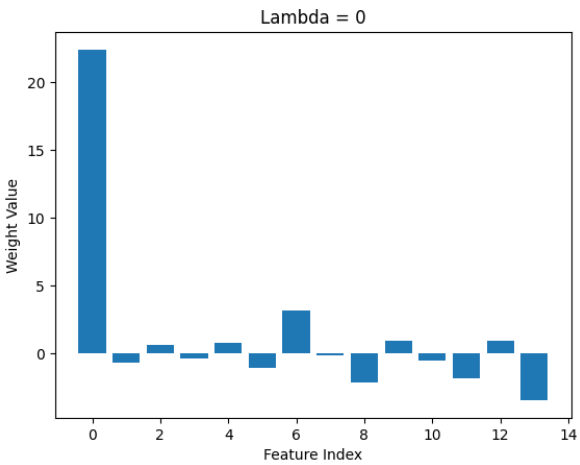
- Histogram of weights history for:
- Plot the cost history for each lambda to demonstrate that your model converged.
- Step length, number of steps, and final cost for the figures

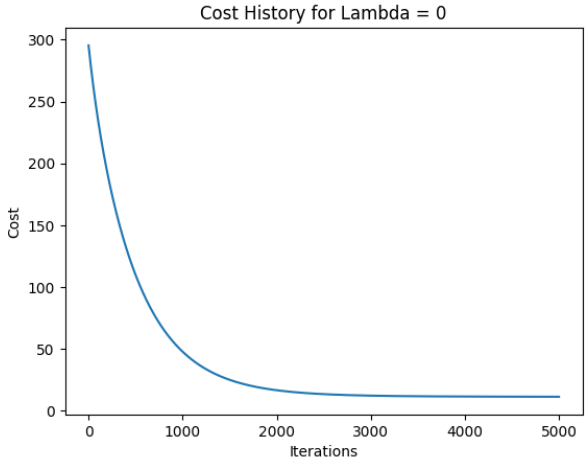
a. $\lambda = 0$

Step Length = 0.001

No. of Steps = 5000

Final cost: 11.347749710083008



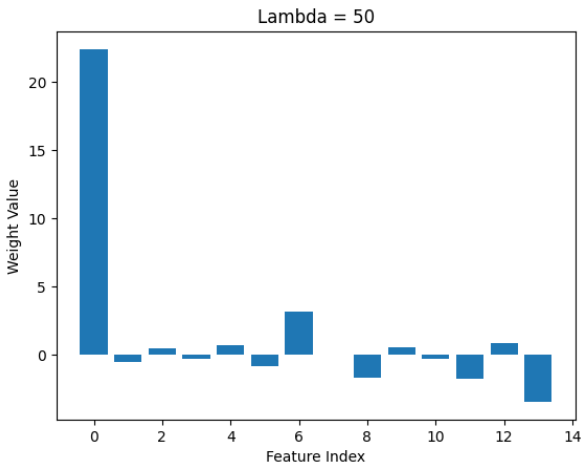



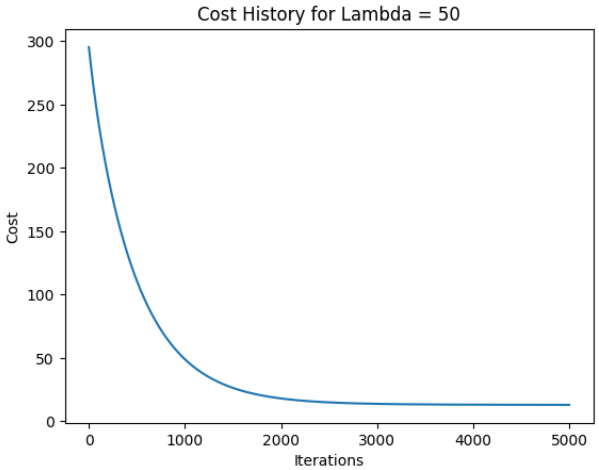

b. $\lambda = 50$

Step Length = 0.001

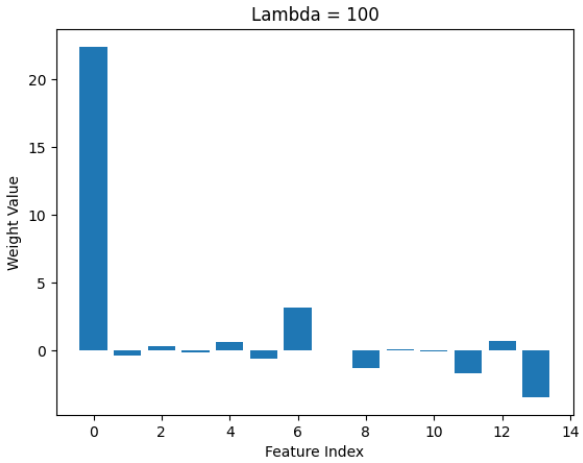
No. of Steps = 5000

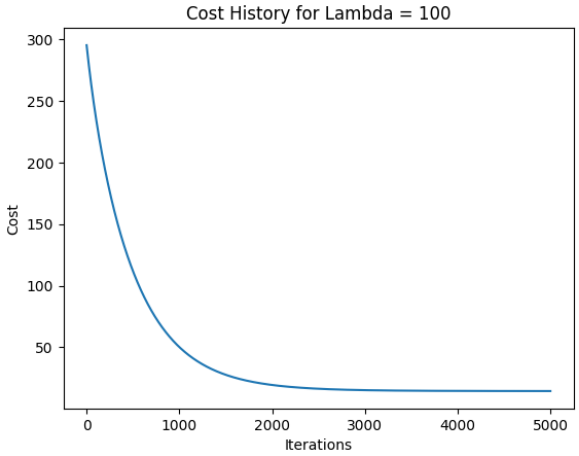
Final cost: 13.034984588623047



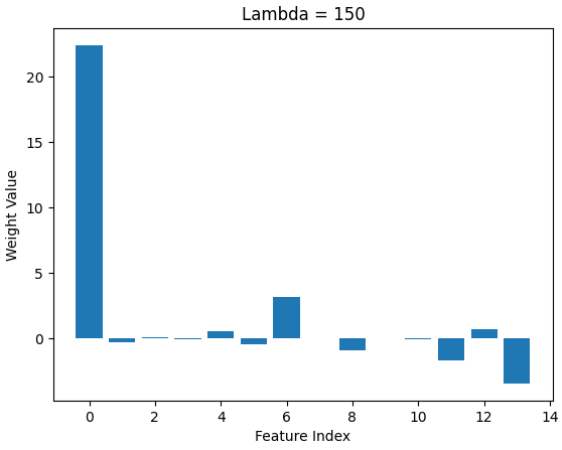



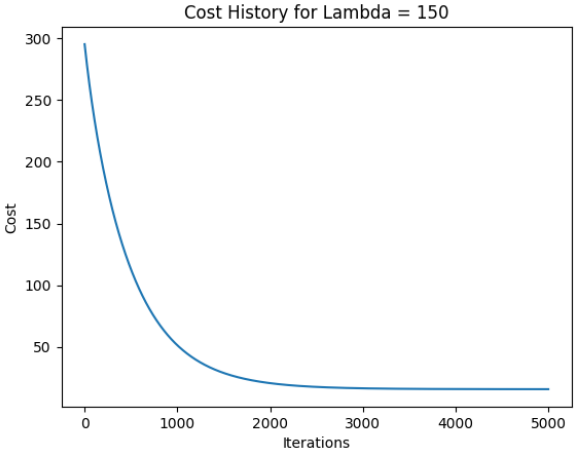
c. $\lambda = 100$
Step Length = 0.001
No. of Steps = 5000
Final cost: 14.451053619384766




d. $\lambda = 150$
Step Length = 0.001
No. of Steps = 5000
Final cost: 15.66240119934082




- Explanation of what you see (and whether or not you can recreate the plots)

- The histograms offer a visual representation of the final weights obtained after the L1-regularized linear regression.
- The line plot illustrates the chronological progression of the cost function during the optimization process.
- An observation from the results indicates that with an increase in lambda, features with minimal impact tend to approach or become exactly zero. This implies that raising the lambda value can highlight the most crucial features, potentially emphasizing the exclusion of less significant features. However, an excessively high lambda might lead to the omission of features that, while less prominent, still contribute meaningfully to the model.