Homework 7

Jian Huang

2023/11/4

Task 1: Comparing standard normalization to PCA-sphering on MNIST

1. Cost history plots during training of using **original data**, **standard normalized data**, and **PCA-spehered data**.

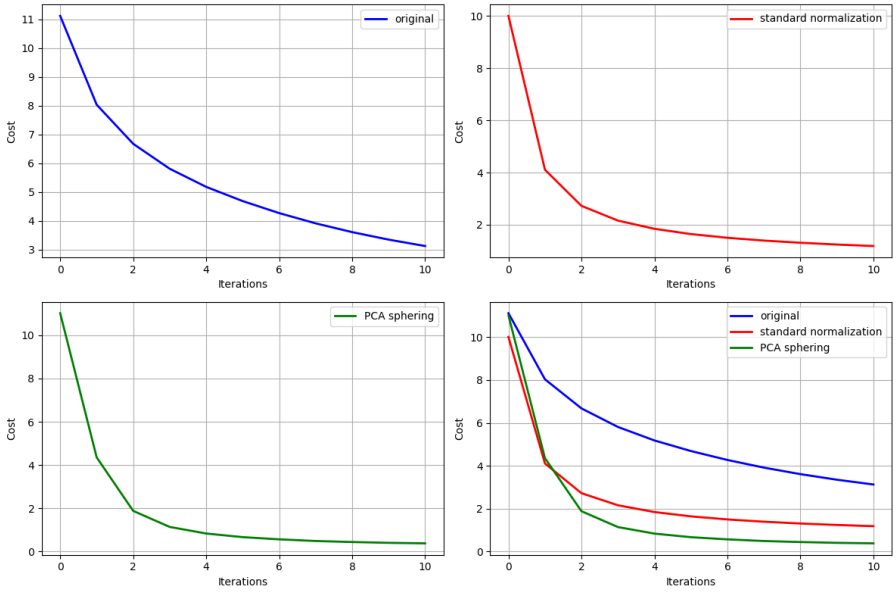

Figure 1: Cost history plots (individually and all together) over 10 iterations

2. Plot of the accuracies over the 10 iterations

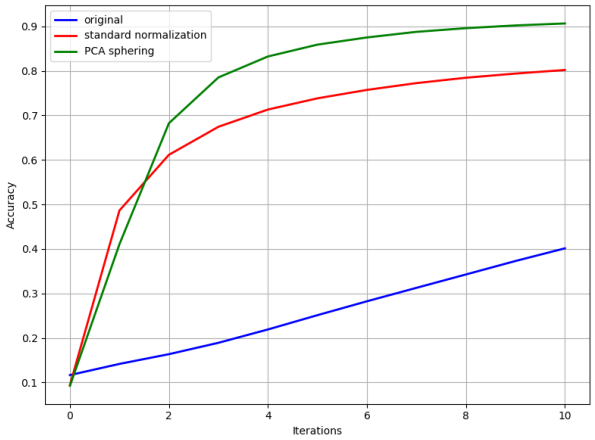

Figure 2: Accuracy over 10 iterations

3. Choice of gamma (the exponent of learning rate α) for **original data**, **standard normalized data**, and **PCA-spehered data**

Data	gamma
original data	-5
standard normalized data	0
PCA-spehered data	1

Table 1: gamma values during training

4. Explanation of what I see

After applying ten iterations of trainings using data engineered differently, clearly traing using the PCA-sphered data produced the lowest cost and also the highest accuracy. This means PCA-sphered data is the most benefical feature engineering method in the given multiclass classification problem.

Task 2: Exploring predictors of housing prices

1. Histogram of weights for $\lambda = 0, 50, 100, 150$.

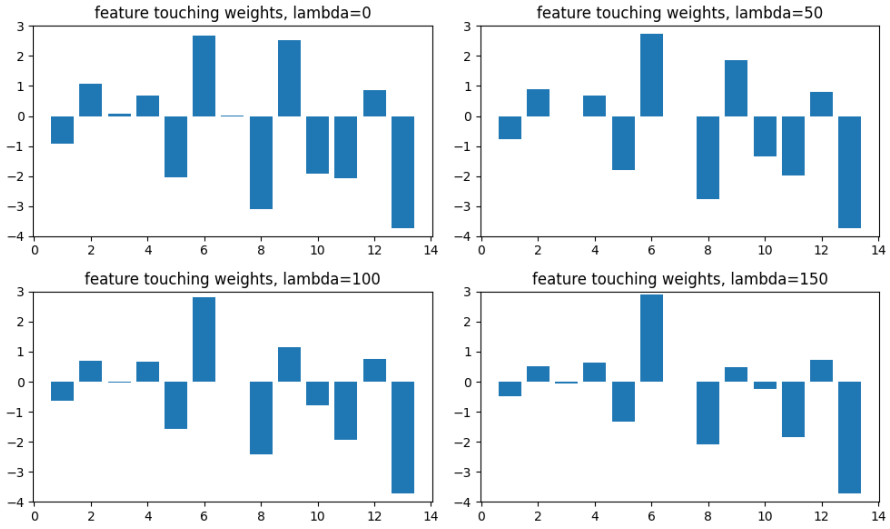

Figure 3: Feature-touching weights for trainings of using different λ

2. Step length, number of steps, and final cost.

λ	step length	number of steps (iterations)	final cost
0	0.1	400	22.778
50	0.1	400	25.066
100	0.1	400	26.652
150	0.1	400	28.176

Table 2: Hyperparameters

3. Cost history plots.

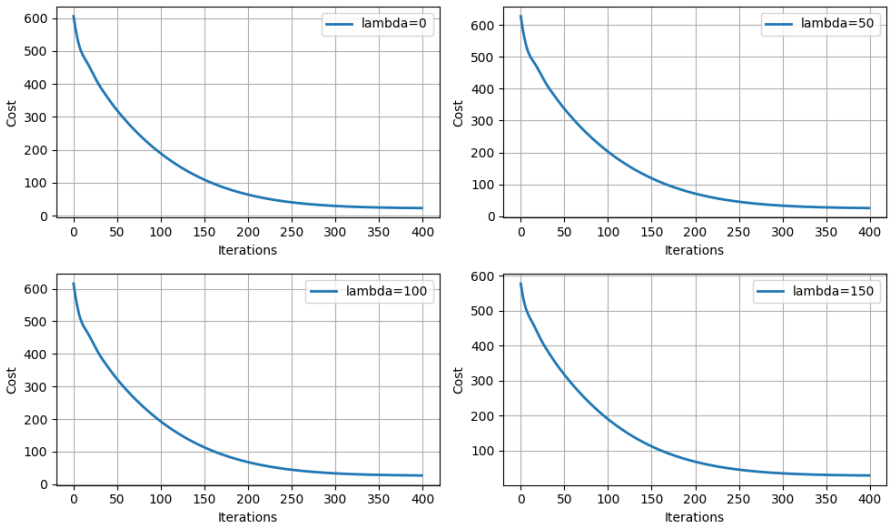

Figure 4: Cost history for each λ

4. Explanation of what I see

Based on Figure 3, there are some weights whose "bars" (absoulte values) are consistently large, such as w6, w8, w13, meaning that they corresponds to the most important features. However, some weights are diminishing as λ increases, such as w5, w9, w10, meaning that the corresponding features are relatively unimportant.