Homwork Assignment 1

Salil Kanetkar, Sanjay Krishna Gouda

October 4, 2017

1 Summary

- Performed simple split, 10 Fold Cross Validation and Leave One Out Cross Validation (LOOCV) on given training set to find optimal λ value that regularizes the model well.
- For the optimal λ *values* (*say* λ^*) found using cross validation, we found the RMS cost on the test set as follows :
  - $\lambda^* = 0.008$
  - kfoldcost = 0.38838
  - simplecost = 0.39420
  - leaveonecost = 0.38979

  Where kfoldcost, simplecost and leaveonecost are the RMS error values on test set found using the simple split, 10Fold CV and LOOCV respectively.
- We then changed the model this time not regularizing the bias term and found the RMS error values on the testset as follows:
  - $\lambda^* = 0.0593$
  - kfoldcost = 0.37906
  - simplecost = 0.39348
  - leaveonecost = 0.37892

2 Procedure

- Using octave, loaded the training data first and extracted the inputs and labels from the first and second column of train.txt into different vectors.
- We then converted the input vector to a matrix X_{10x20} where the n^{th} column of X is $(1, x_n, x_n^2,_n x^3, ..x_n^9)^T$.

2.1 Descretization of λ

- To test out different ranges of λ, we used an array that contained λ values ranging from $\log(\lambda)$ = -18 to $\log(\lambda)$ = 4. We then plotted the loss using simple split on training set in $\log(\lambda)$ space to find a region where optimal λ could lie in.
- In this neighborhood, we constructed an array with a finer range of λ values. It was constructed as follows : $lambda(i) = 0 + (0 : 5000 - i) * 0.0002$. A total of 5000 values equally spaced in the range of 0 and 0.09980 which we used during the analysis of different CV methods.
- For each of these λ values, we performed simple split, kFold and LOOCV on the training set.
- For the initial simple split method we used 75% of training set for training remaining validation. That is, 15 points for training and the remaining 5 points for validation.

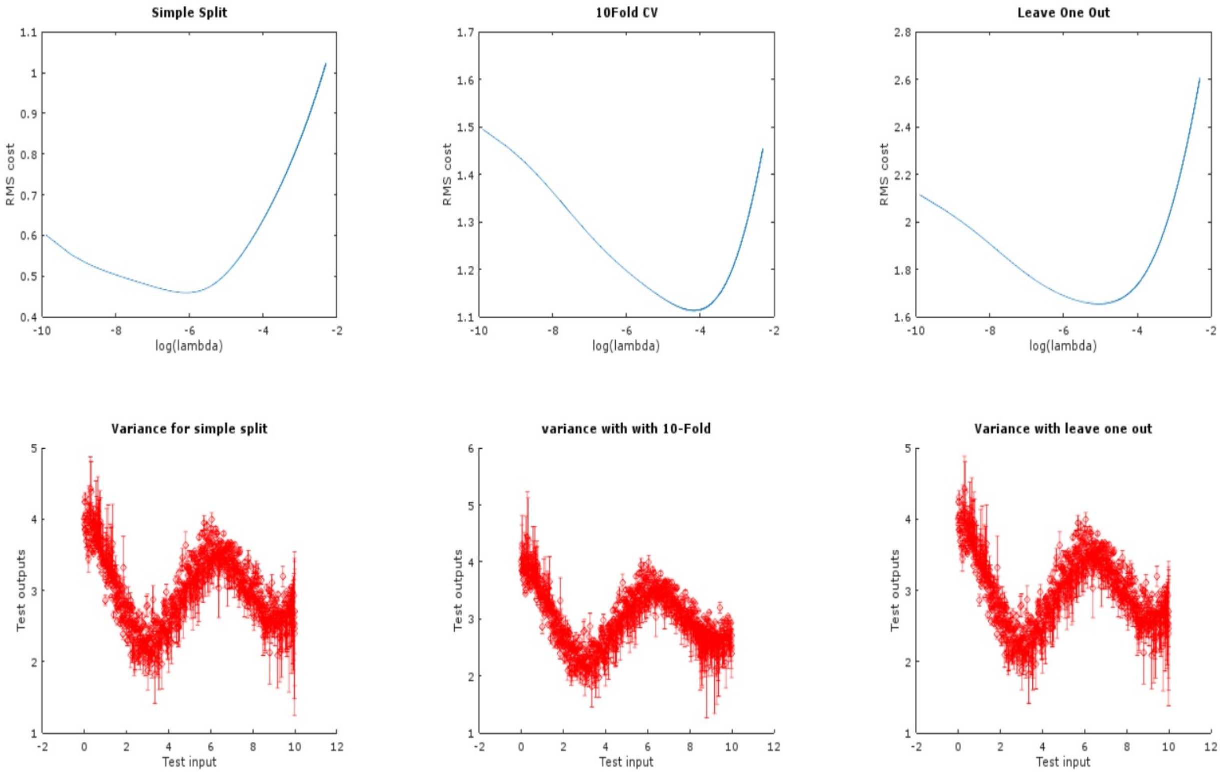

Figure 1: cost vs $\log(\lambda)$ and Variance plots.

- For performing kfold cross validation and LOOCV, we looped over appropriate indices for training and validation.
- Using the min function we found the λ^* and the lowest cost. Using this λ^* we found the corresponding weights matrix using : $W* \;=\; (XX^T + \lambda I)^{-1}Xt$. Where X is the matrix that was described above and t is the vector that contains the labels (second column from train.txt).
- We repeated the same procedure using 10Fold CV and LOOCV.
- Once we had the λ^* for all the validation methods, we constructed corresponding W^* and plotted the cost vs $\log(\lambda)$ on the training data set.
- We then used these weight matrices (W^*) in each case to test the model on the test set.
- To exclude the bias term from regularization, we modified the equation:

  $$W^* = argmin_w(\sum_n \sum_{i=0to9} (w_i x_n - t_n)^2 + \lambda \sum_{i=0to9} w_i^2)$$

  or in the vector form:

  $$W^* = argmin_w(||X^T W - t||^2 + \lambda ||W||^2)$$

  to

  $$W^* = argmin_w(\sum_n \sum_{i=0to9} (w_i x_n - t_n)^2 + \lambda \sum_{i=1to9} w_i^2)$$

  or in the vector form:

  $$W^* = argmin_w(||X^T W - t||^2 + \lambda ||W(2:10)||^2)$$

Conclusion

- We found the difference in cost of including or excluding the bias term in regularization to be insignificant.
- LOOCV for this dataset is more consistent in a sense that the λ^* that we get after each run of the algorithm remains the same. Where as in 10Fold CV, the splits of the data are random and hence the effective λ^* changes every time.
- Because the test error is smaller for 10FoldCV and also it can be seen from the plots that it has better variance bars for the losses as compared to LOOCV and simple split. Thus in this sense, we can say that 10FoldCV is better.