CS584 Final Project: Report

Lida Kuang
CWID: A20403666

August 12th , 2018

Abstract

This is a report for final project in CS 584. The report is at least 4-6 pages long. It contains the problems I have solved or tried to solve, the algorithms employing to solve problems, and the results. It also contains analyzed results and the evaluation of the algorithms.

1. Problem Statement

- Topic: Predicting the concrete strength by using Neural Network and Kernel Ridge Regression
- Data processing
  - Data normalization
  - Data split for training and testing
- Training----Neural network and Kernel ridge regression
- Classify
- Evaluation
- Modification

2. Proposed Solution

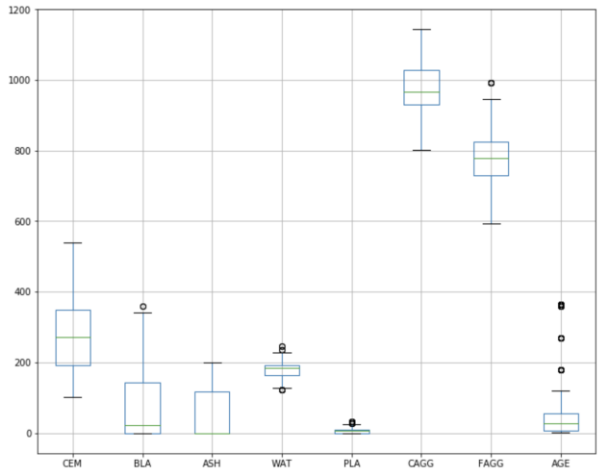
- Training----Neural network and Kernel ridge regression
  - Build neural network module by TensorFlow.
  - Apply SGD to compute parameter.
  - Derive L2 linear least square of ridge regression, and combine with kernel trick to get formula of optimal parameter.
- Evaluation
  - MSE
  - R2-score
- Modification
  - For neural network, change number of layers and number of node to compare the results
  - For kernel ridge regression, change the value of parameter and degree to compare the results

3. Implementation Details

- Data Processing
  - Source: http://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength
  - Description:

    This dataset contains 1030 instances, 8 features and 1 class.

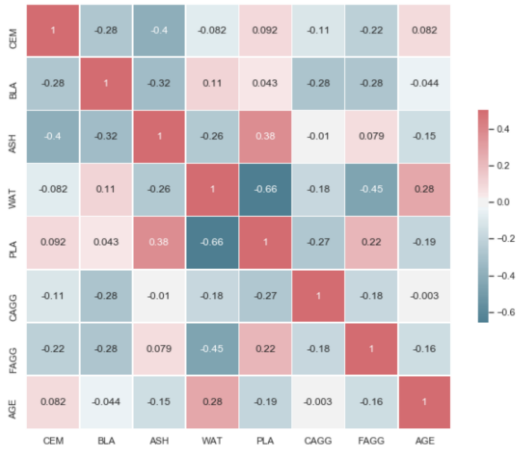
    Data type is quantitive.
  - Distribution:



    Because of the huge difference in range of features, in order to eliminate the influence of the scaling of features, data normalization is necessary.
  - Normalization:

$$z = \frac{x - \bar{x}}{S}$$

  - Feature correlation:



    According to the correlation matrix above, features are almost independent, so no added feature to original dataset

- Training
  - Neural Network
    * Build model by keras.Sequential
    * (SGD) Set epoch value to 500, for every 100 epoch, check the improvement
    * Plot training loss and validation loss
    * Define functions to compute MSE, R2-score and MAE
  - Kernel Ridge Regression
    * Derive optimal parameter formula of Kernel Ridge Regression:

$$C(w) = \frac{1}{2}\sum_i (y_i - W^T\phi)^2 + \frac{1}{2}\lambda\|W^T W\|$$

$$optimal\ w = \sum_i \alpha_i \phi(x_i)$$

$$C(\alpha) = \frac{1}{2}\sum_i (y_i - \sum_j \alpha_j \underbrace{\phi(x_j)\phi(x_i)}_{K(x_i, x_j)})^2 + \frac{1}{2}\lambda \sum_i \sum_j \alpha_i \alpha_j \underbrace{\phi(x_j)\phi(x_i)}_{K(x_i, x_j)}$$

$$C(\alpha) = \frac{1}{2}\lambda I \boldsymbol{\alpha}^T K \boldsymbol{\alpha} + \frac{1}{2}(Y - \boldsymbol{\alpha}^T K)^2$$

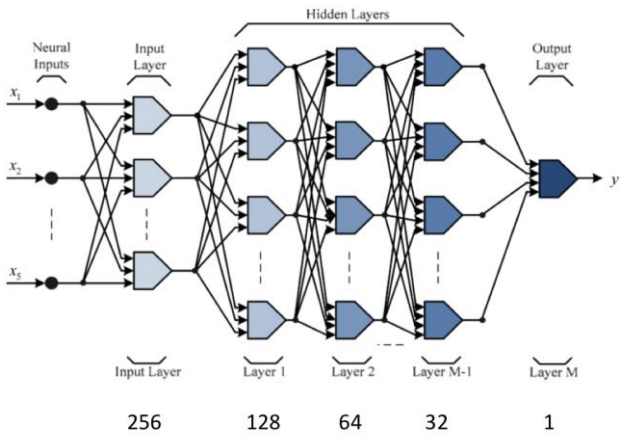
    * Call KernelRidge function to get the model of kernel ridge regression
    * Change parameter and compare results

4. Results and Discussion

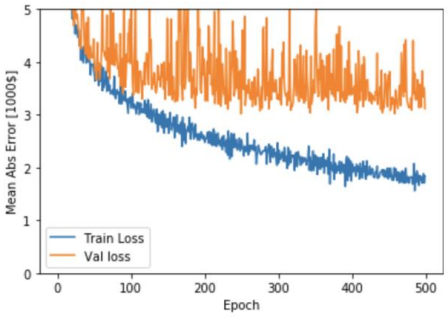
- Basic linear model
  - Use basic linear model to do initial prediction:

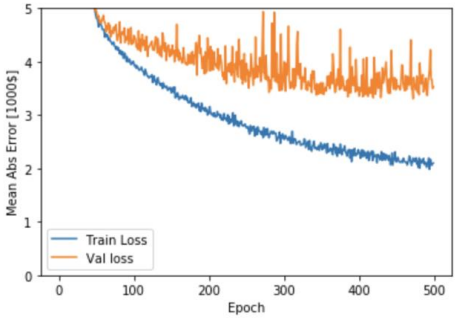
```
R2 Score: 0.5210125654324624
R2 Score: 0.5404158815770888
R2 Score: 0.6627321309593508
R2 Score: 0.6867785806136337
R2 Score: 0.619768021051541
Average R2 Score 0.6061414359268154
```

    This R2-score is supposed to be the lowest among that of Neural Network and KRR
- Neural Network
  - Model: 5 layer --- 4 hidden layer and 1 output layer



- Compare the result of layer number:



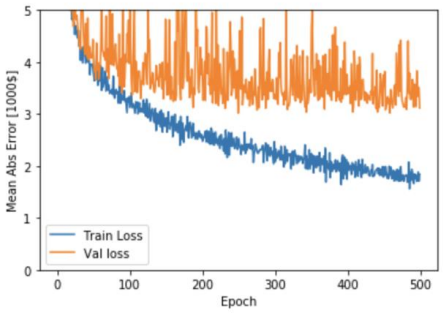


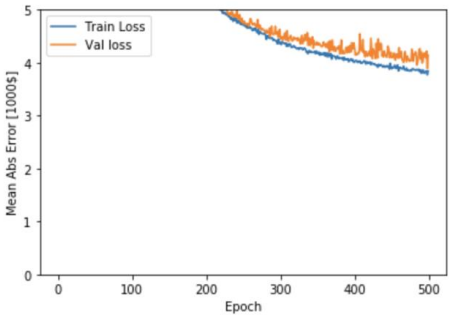
- 5-layer-NN
- R2 score: 0.91180
- MSE: 23.10260

- 3-layer-NN
- R2 score: 0.84249
- MSE: 41.25606

  * Orange line represents validation loss, blue line represents training loss, both of them are converge to some point while iteration has done.
  * Orange line(val loss) is not as smooth as blue line, because the size of testing dataset is smaller that training, so it should be much more sensitive to noise.
  * After reduce the number of layers, R2 score goes low and MSE goes high, which means 3-layer model is not as good 5-layer one

- Compare the result of node number:



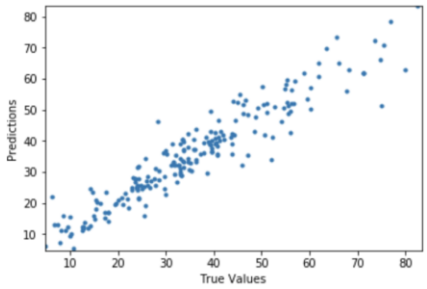


- 5-layer-NN
- R2 score: 0.91180
- MSE: 23.10260

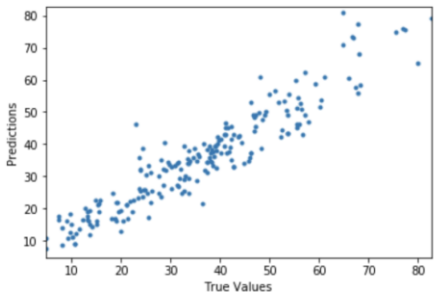
- 5-layer-with less nodes
- R2 score: 0.84631
- MSE: 40.25470

  * Left figure shows model containing 43985 nodes, right figure shows model containing 622 nodes.
  * After reduce the number of nodes, R2 score goes low and MSE goes high, which means less node model is not as original one

- Kernel Ridge Regression:
    - Compare different parameter under the same degree:
        - Degree=4 , among α=1, 0.1, 0.01, 0.001
          Best estimator: $\alpha = 0.01$
          R2 score: 0.9033045893841107



Degree=4 , among α=1, 0.1, 0.001
Best estimator: $\alpha = 0.1$
R2 score: 0.8810557818603341



* Keep degree to 4, after deleting the best estimator from estimators set in model one(left), chose the second-best estimator among remain estimators, R2-score goes low, which means the model is not as good as before.

5. Reference

- Max Welling, “Kernel ridge Regression,” Department of Computer Science University of Toronto 10 King’s College Road Toronto, Canada.
- Predicting the Compressive Strength of Concrete using Neural Network and Kernel Ridge Regression
- An introduction to statistical learning with applications in R
- TF using with 2 hides NN: https://www.tensorflow.org/tutorials/keras/basic_regression
- Feedforward Neural Network: https://en.wikipedia.org/wiki/Feedforward_neural_network
- CG: http://matlab.izmiran.ru/help/toolbox/nnet/backpr59.html
- A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, MARTIN FODSLETTE MEILLER : http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.294.6699&rep=rep1&type=pdf