MACHINE LEARNING APPLICATIONS TO PROCESS ENGINEERING

Poly-Methyl Methacrylate Reactor Unit

Ananya Bansal
Lakshya Sharma

Indian Institute of Technology, Ropar

May 11, 2024

CONTENTS

Introduction:

What is PMMA, what does the process encompass, and what is our aim for this model

Introduction to the process

The PMMA production process involves a **free-radical polymerization** of **methyl methacrylate (MMA)** using a **Continuous Stirred-Tank Reactor (CSTR)** configuration. **Azo-bis-isobutyronitrile (AIBN)** serves as the **initiator** for the polymerization reaction, while **toluene** functions as the solvent.

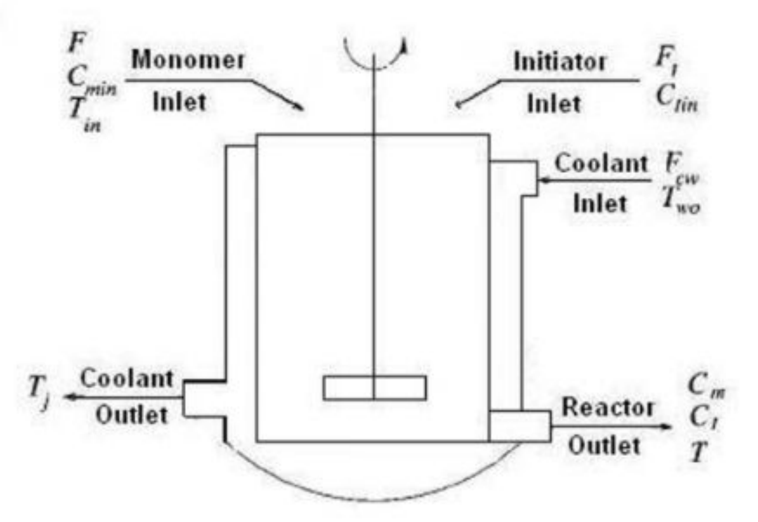


Fig 1: MMA polymerization reactor flow sheet

Process Description:

- MMA, AIBN, and toluene are continuously fed into the CSTR.
- AIBN **decomposes** upon heating, releasing free radicals that **initiate** MMA polymerization.
- MMA monomers react with the **free radicals**, forming PMMA chains.
- Heat generated from the exothermic reaction is controlled by a cooling jacket.
- PMMA **chains** grow in **length**, forming desired polymer product.
- Polymer solution is continuously removed from reactor for further processing.

Objectives:

- Optimizing the process of MMA to produce PMMA efficiently.
- To maintain a **target monomer concentration** in the reactor for **consistent product quality and reaction efficiency.**
- Developing **predictive models** for monomer concentration using **multivariate regression** techniques.
- To improve operation stability, efficiency, and safety and economic viability of the PMMA process.

Implementation:

Understanding and selection of features, Developing Predictive models based on various approximation methods and Testing, Presentation of Results

Implementation

The process uses 7 input variables, as can be seen from fig 1, two slides before:

- **Temperature (T):** Temperature at the reactor outlet.
- **Jacket Temperature (Tj):** Temperature of the coolant outlet.
- **Coolant water flow rate (Fcw):** Rate at which water flows through the system for temperature regulation.
- **Monomer inlet Flow rate (F):** Rate at which monomer is introduced into the reactor.
- **Coolant inlet temperature of water (Two):** Initial temperature of the coolant entering the system.
- **Inlet temperature (Tin):** Initial temperature of the input material entering the reactor.

And target variable as

- **Monomer concentration (Cm):** The concentration or amount of the monomer at the reactor outlet.

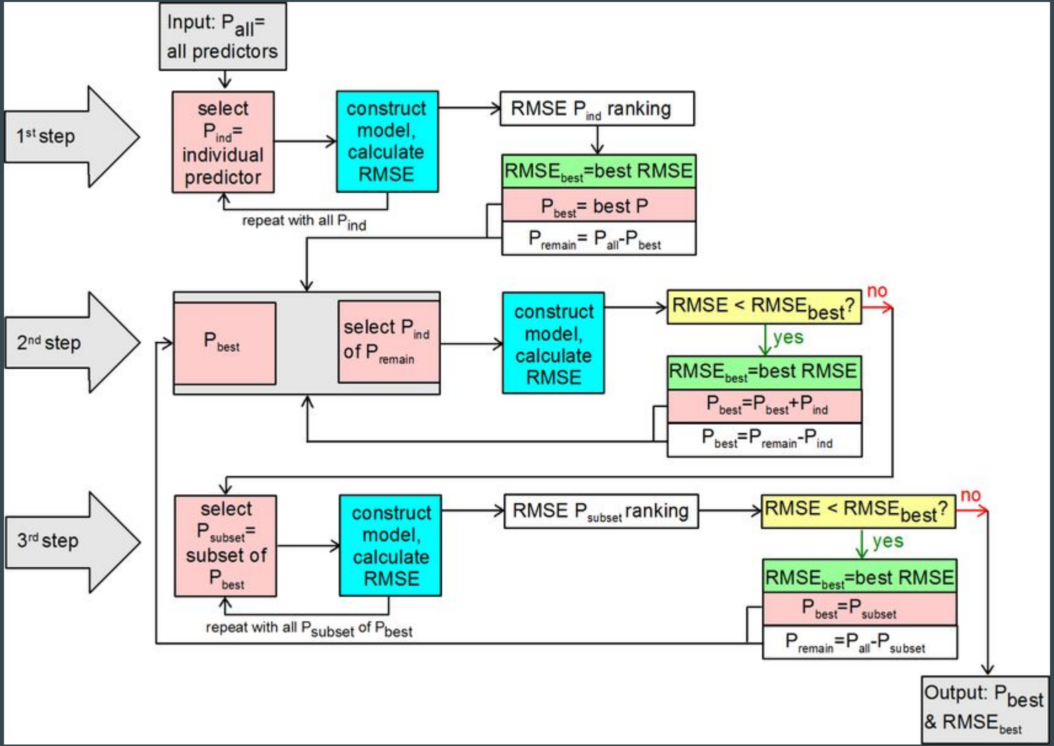
For modelling, **Fcw** and **Two** were directly dropped as they remain constant for the entire data set.

Parameters remaining, with index 0 to 4:

0: Initiator Concentration(CI)
1: Temperature (T)
2: Jacket Temperature (Tj)
3: Monomer inlet Flowrate (F)
4: Inlet temperature of Feed (Tin)

	Initiator Concentration(CI)	Temperature (T)	Jacket Temperature (T)	Coolant water flowrate (Fcw)	Monomer inlet Flowrate (F)	Coolant inlet temp.water (Two)	Inlet temparature of Feed (Tin)	Target variable (Monomer Concentration)
count	6000.000000	6000.000000	6000.000000	6.000000e+03	6000.000000	6000.0	6000.000000	6000.000000
mean	0.025075	349.430084	331.626363	1.588000e-01	1.000060	293.2	350.000100	6.039932
std	0.000238	3.545478	3.322300	2.775789e-17	0.009995	0.0	0.079624	0.012250

Linear: Subset Selection



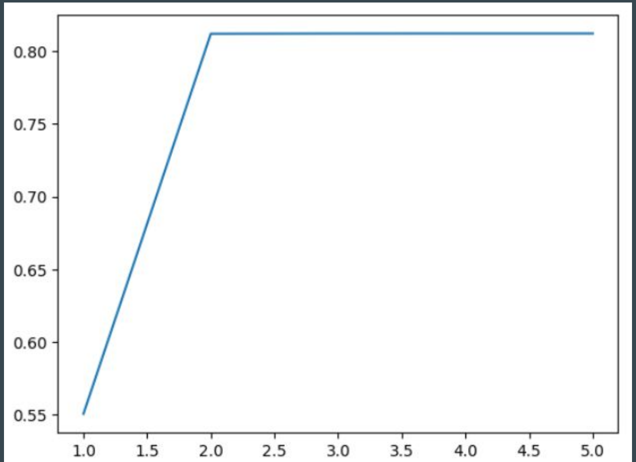
- ➔ Define response variable y and matrix of predictors X .
- ➔ Divide data into training, testing data.
- ➔ Determine number of predictors **p** and define all possible subsets **S(p, k)**
- ➔ For each subset S(p, k):

  1. **Fit linear regression model y = XβS + ε**

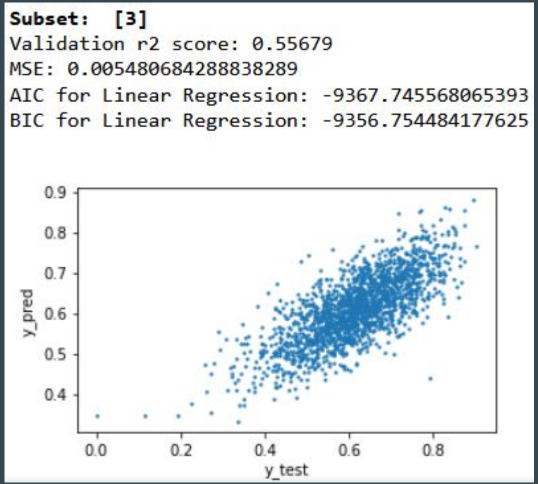
  2. Calculate SSE(k), R2(k), AIC(k), BIC(k) for performance evaluation.
- ➔ Select subset S(p, k) that gives best result
- ➔ Interpret coefficients of selected model and hence decide which features to eliminate further for optimal results

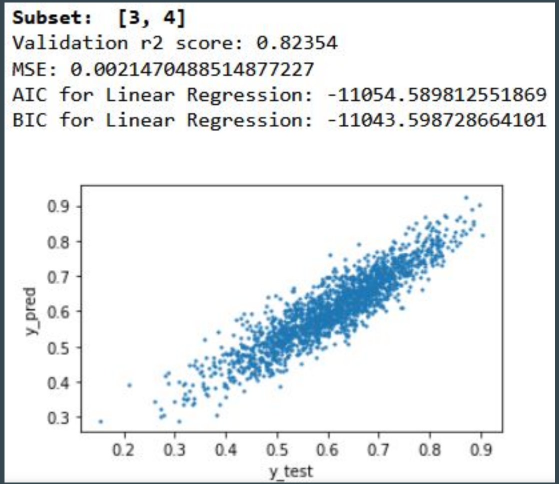
Optimal parameter subsets and their obtained R2 values:

[3]: 0.55679 [3, 4]: 0.82354 [2, 3, 4]: 0.83216 [0, 1, 3, 4]: 0.82306 [0, 1, 2, 3, 4]: 0.86836

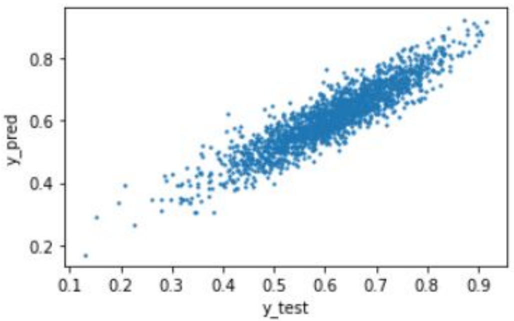


R2 vs number of features

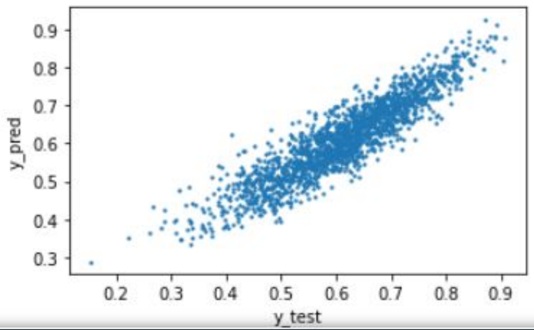




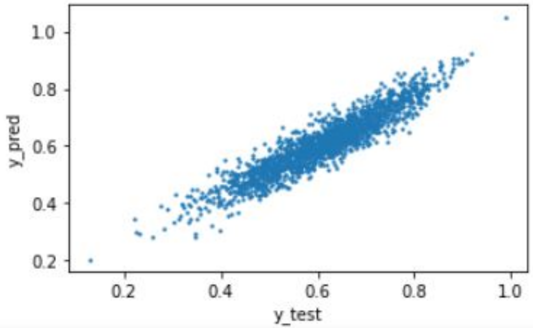
```
Subset:  [2, 3, 4]
Validation r2 score: 0.83216
MSE: 0.00207862290385172
AIC for Linear Regression: -11112.889405976475
BIC for Linear Regression: -11101.898322088708
```




```
Subset:  [1, 2, 3, 4]
Validation r2 score: 0.82306
MSE: 0.002157082708524253
AIC for Linear Regression: -11046.19742252762
BIC for Linear Regression: -11035.206338639853
```




```
Subset:  [0, 1, 2, 3, 4]
Validation r2 score: 0.86836
MSE: 0.0017092839361859502
AIC for Linear Regression: -11465.02534448589
BIC for Linear Regression: -11454.03426059812
```



Hence, the features selected would be: [3,4] or [2,3,4] (Refer to Slide 8)

Linear: Lasso Regression

It works by introducing a bias term, ie the absolute value of the slope is added as a penalty term.

Define response variable y and matrix of predictors X .

→ Standardize the predictor variables X

→ Initialize $\beta = 0$, λ (lambda) = large value

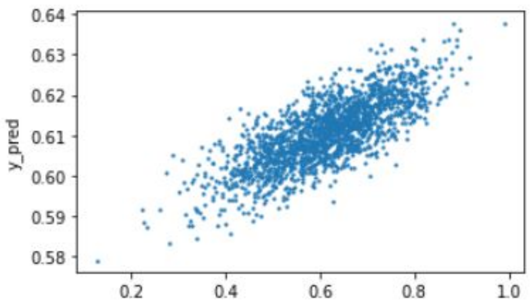
→ For a decreasing sequence of λ values:
1. Solve:
$\min_{\beta} ||y - X\beta||$^2
subject to: $\Sigma\ |\beta j| \leq t$
2. Update β

→ Select the λ value and corresponding β coefficients

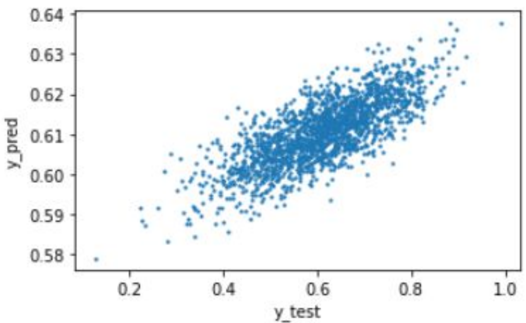
→ Interpret coefficients of selected model and hence decide which features to eliminate further for optimal results

Optimal parameters and their obtained values:

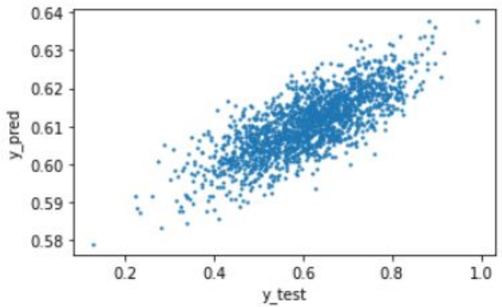
Subset: [3]
Validation r2 score: 0.10361
MSE: 0.011639509099367238
AIC for Linear Regression: -8004.030020091702
BIC for Linear Regression: -7971.056768428396



Subset: [3, 4]
Validation r2 score: 0.10361
MSE: 0.011639509099367238
AIC for Linear Regression: -8004.030020091702
BIC for Linear Regression: -7971.056768428396

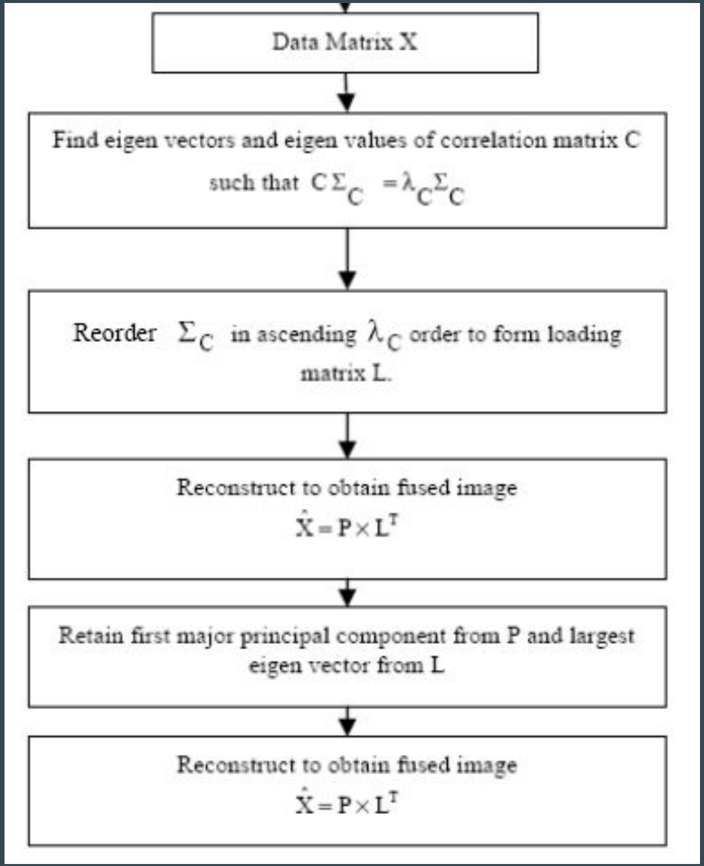


Subset: [0, 1, 2, 3, 4]
Validation r2 score: 0.10361
MSE: 0.011639509099367238
AIC for Linear Regression: -8004.030020091702
BIC for Linear Regression: -7971.056768428396



The model performs poorly for the optimal subsets if a regularisation term is introduced.

Linear: Principal Component Analysis (PCA/PCR)



Define data matrix X with n observations and p predictors

Standardize columns of X

Compute covariance matrix C = X'X

Find eigenvalues λ, and eigenvectors V of the covariance matrix
$CV = \lambda V$

Sort eigenvectors V in descending order of their eigenvalues λ

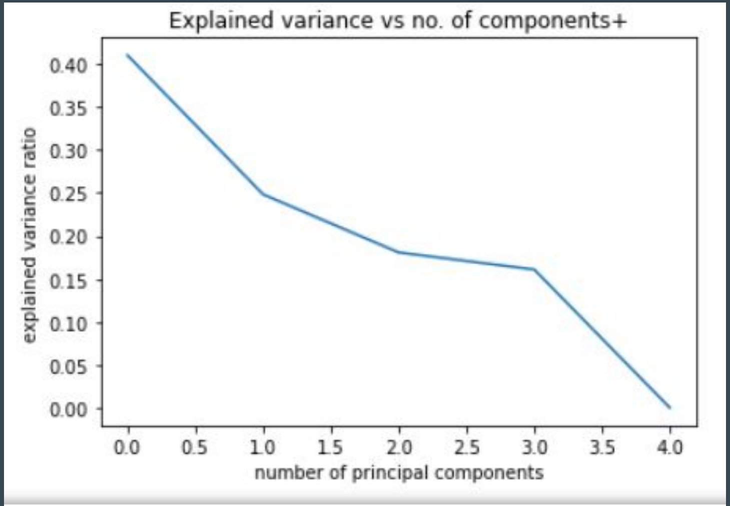
Select the number of principal components (k)to retain based on proportion of variance explained or other criteria

Construct principal component scores:
$T = XV_k$
Regress response variable y on principal component scores T
$y = Tq + e$

Obtain regression coefficients q for the principal components

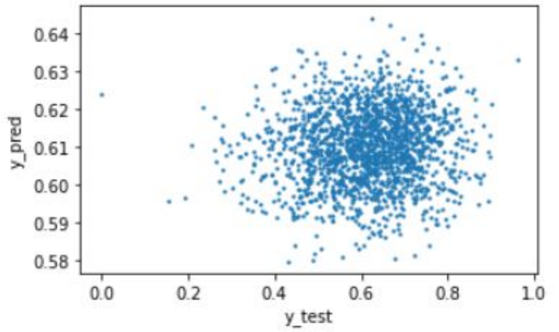
Calculate coefficients for original predictors
$\beta = V_k * q$
And then interpret β

Explained variance vs Number of Components

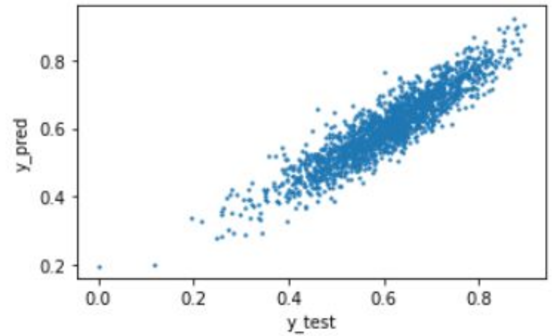


- Elbow at 2 and then 4.
- PCR performs worse than traditional regression for n_components=2 and has comparable performance for n_components=4

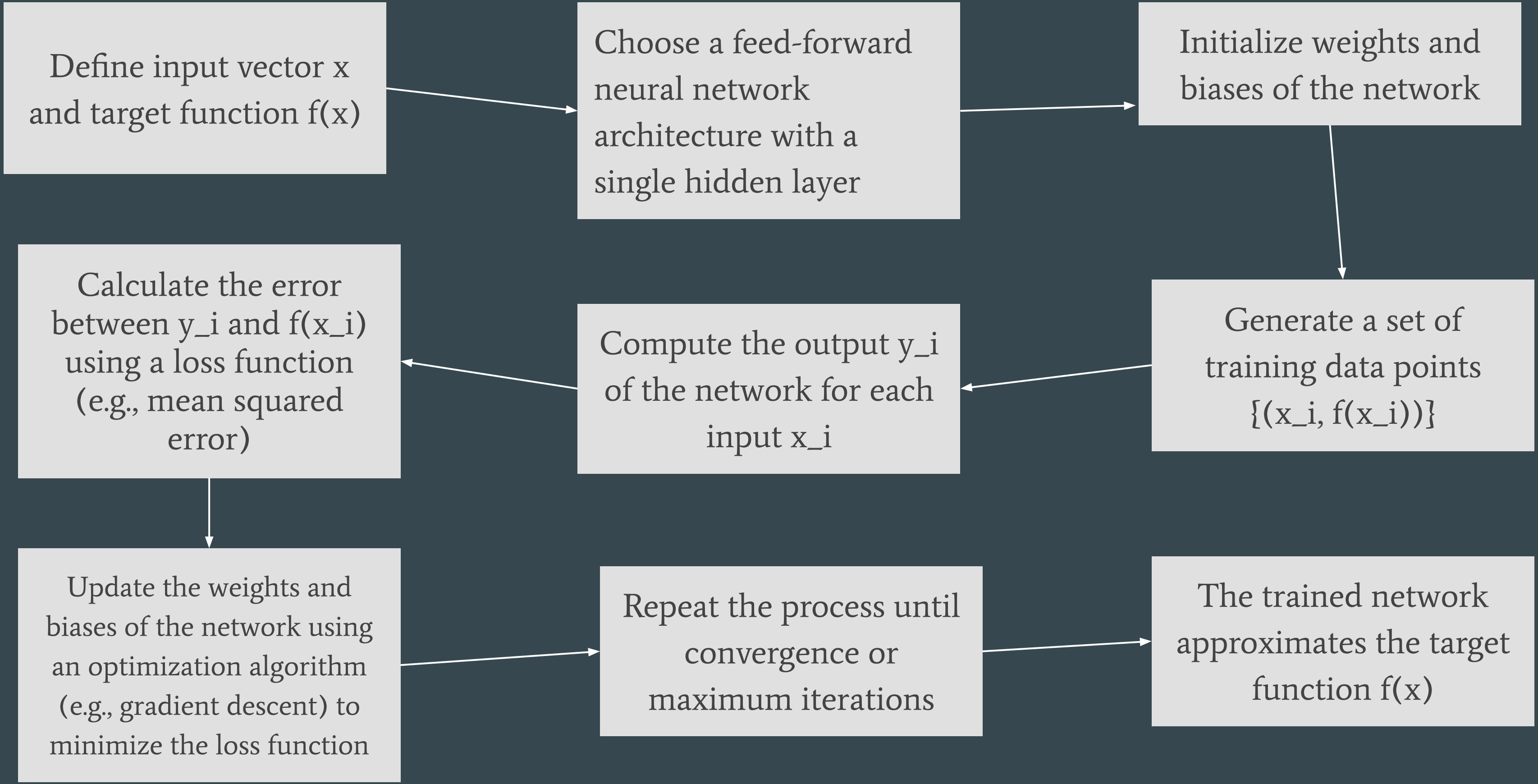
```
Number_of_components 2
Validation r2 score: 0.006190855951851426
Validation mse score: 0.012919676601512905
AIC for Linear Regression: -26090.022870701676
BIC for Linear Regression: -26076.623841205255
```




```
Number_of_components 4
Validation r2 score: 0.8393353611752034
Validation mse score: 0.0020590881557835455
AIC for Linear Regression: -37108.95222199654
BIC for Linear Regression: -37095.55319250012
```



Nonlinear: Feed Forward (Shallow Neural Network)



Nonlinear: Feed Forward (Shallow NN)

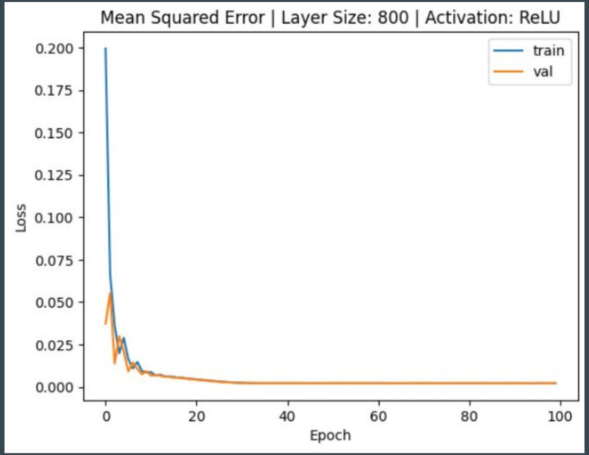
| Effect of Activation Functions

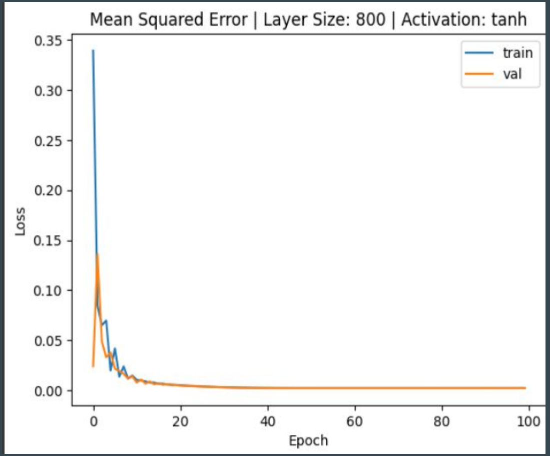
Keeping the number of nodes in hidden layer fixed at 800, we see

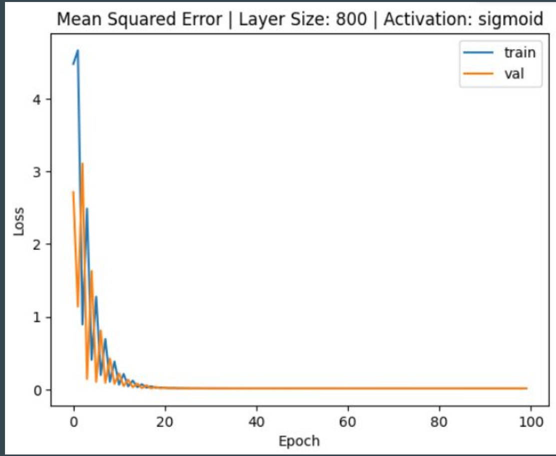
```
Activation Function:  ReLU
Validation r2 score: 0.13207
Total no. of parameters: 644801
MSE: 0.0021650740391680784
AIC for Linear Regression: 1278558.458704211
BIC for Linear Regression: 4822089.399662724
```

```
Activation Function:  tanh
Validation r2 score: 0.14272
Total no. of parameters: 644801
MSE: 0.0021544549523039754
AIC for Linear Regression: 1278549.608482393
BIC for Linear Regression: 4822080.549440905
```

```
Activation Function:  sigmoid
Validation r2 score: 0.9989
Total no. of parameters: 644801
MSE: 0.012021620251940632
AIC for Linear Regression: 1281644.1125872754
BIC for Linear Regression: 4825175.053545787
```







TAKEAWAY: While performance of ReLU and tanh is comparable, tanh shows a better learning curve than Sigmoid.

Nonlinear: Feed Forward (Shallow NN) | Effect of Layer Size

For tanh activation function,

Layer size: 2

```
Activation Function:  tanh
Validation r2 score: 0.83423
Total no. of parameters: 17
MSE: 0.02339032123002743
AIC for Linear Regression: -6725.779336698184
BIC for Linear Regression: -6632.355123652152
```

Layer size: 40

```
Activation Function:  tanh
Validation r2 score: 0.10607
Total no. of parameters: 1841
MSE: 0.002233633071402542
AIC for Linear Regression: -7305.426510529513
BIC for Linear Regression: 2811.866208161404
```

Layer size: 800

```
Activation Function:  tanh
Validation r2 score: 0.14272
Total no. of parameters: 644801
MSE: 0.0021544549523039754
AIC for Linear Regression: 1278549.6084823934
BIC for Linear Regression: 4822080.549440905
```

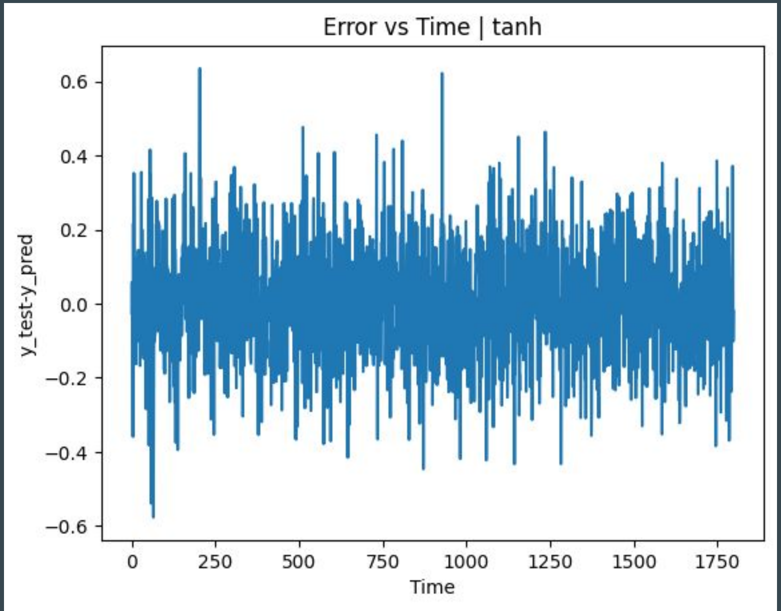
TAKEAWAYS

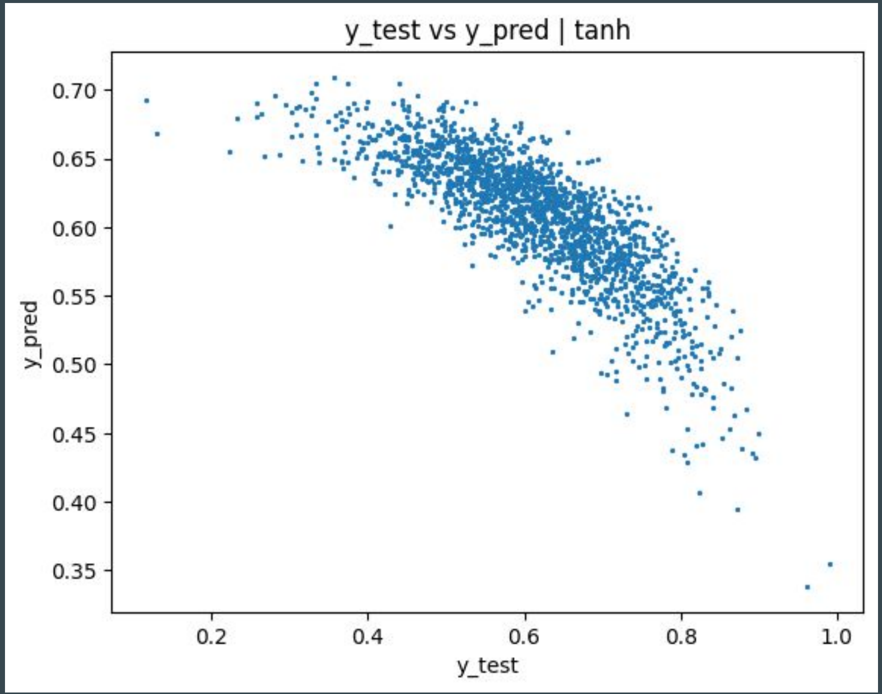
Performance Metric: MSE improves with layer size

BIC being more conservative penalises layer size 40 to be worse than layer size 2 which is underfitting, While AIC considers it better.

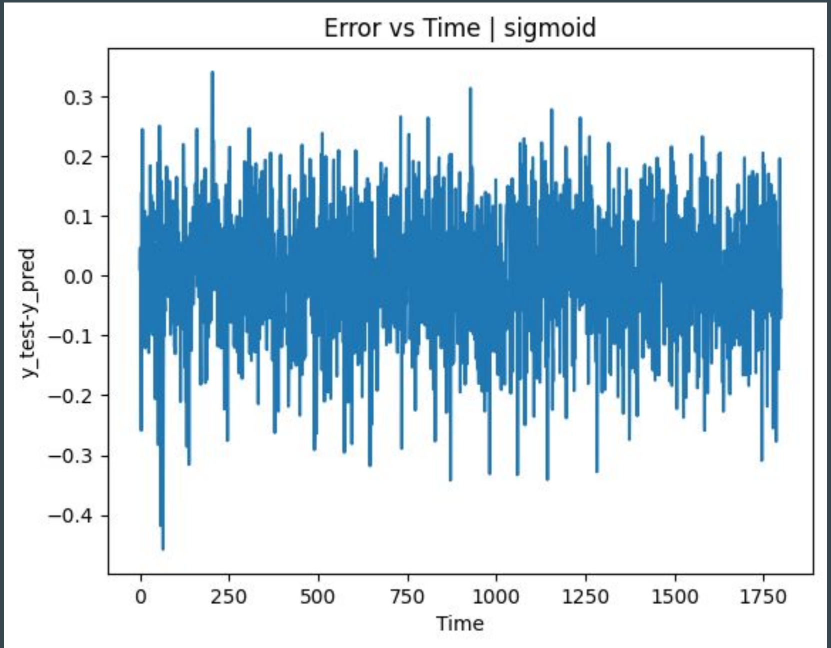
The number of trainable parameters increase exponentially with layer size

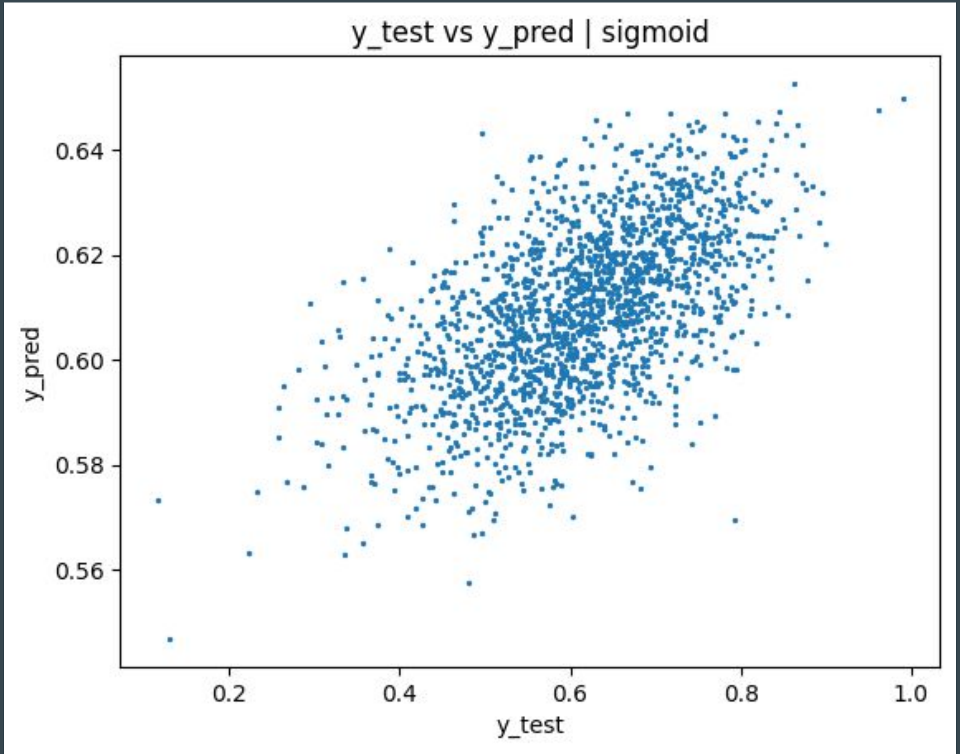
```
Activation Function:  tanh
Validation r2 score: 0.83423
Total no. of parameters: 17
MSE: 0.02339032123002743
AIC for Linear Regression: -6725.779336698184
BIC for Linear Regression: -6632.355123652152
```





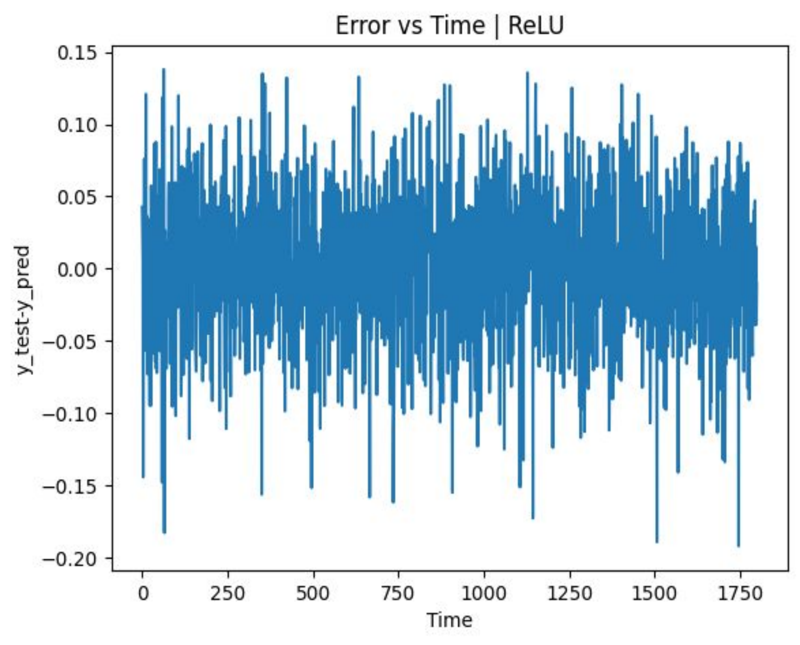
```
Activation Function:  sigmoid
Validation r2 score: 0.98098
Total no. of parameters: 17
MSE: 0.011004929270775952
AIC for Linear Regression: -8082.941583858423
BIC for Linear Regression: -7989.51737081239
```

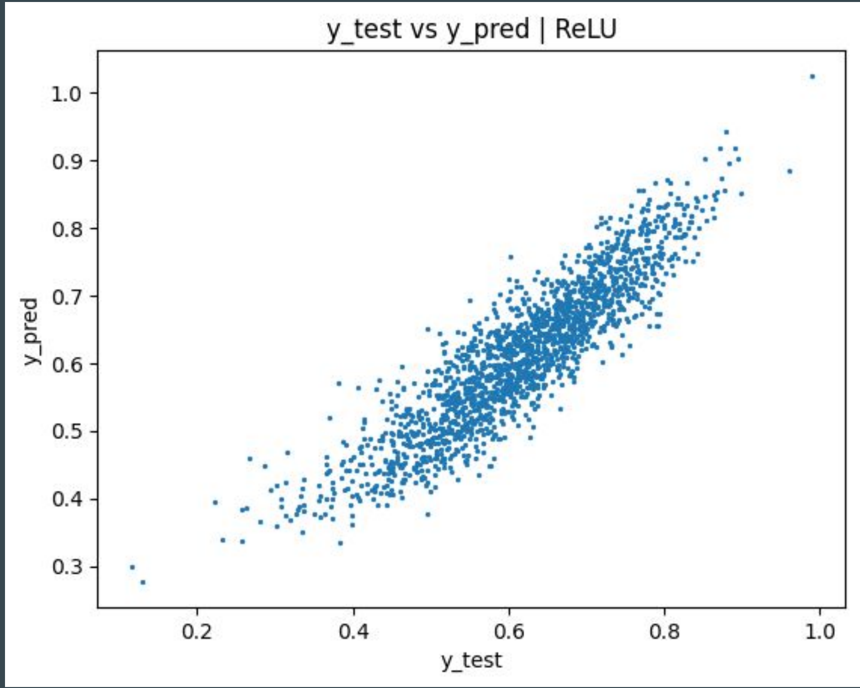




Layer Size: 40

Activation Function: ReLU
Validation r2 score: 0.09309
Total no. of parameters: 1841
MSE: 0.0023713833696311732
AIC for Linear Regression: -7197.707228939935
BIC for Linear Regression: 2919.5854897509817





Activation Function: tanh

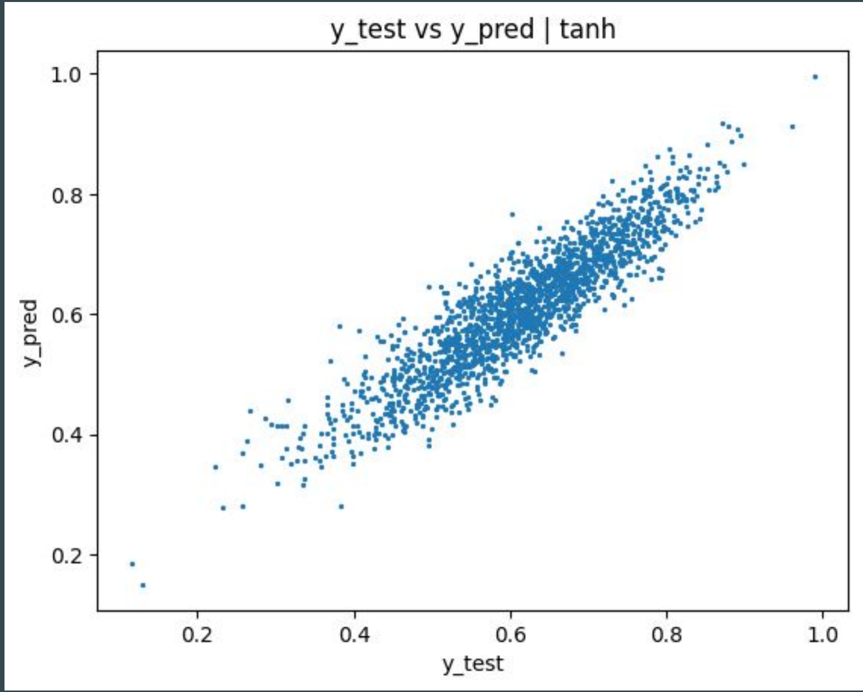
Validation r2 score: 0.10607

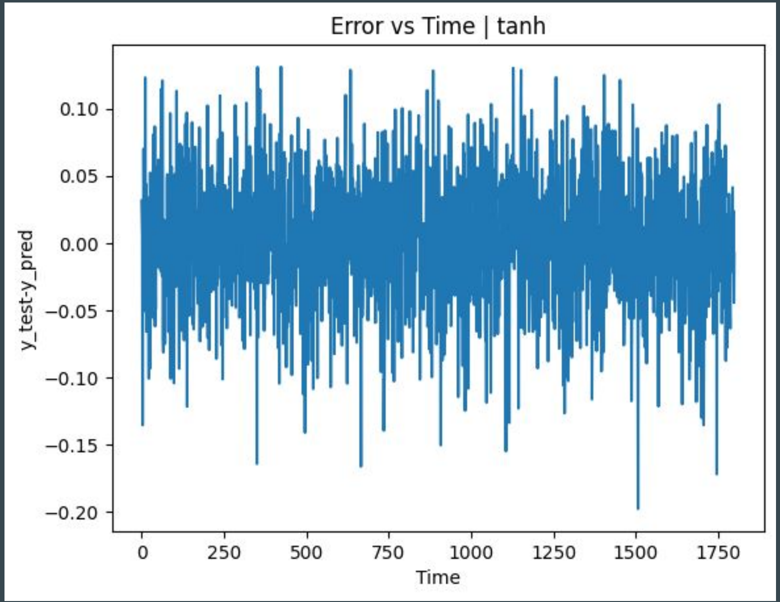
Total no. of parameters: 1841

MSE: 0.0022336330714O2542

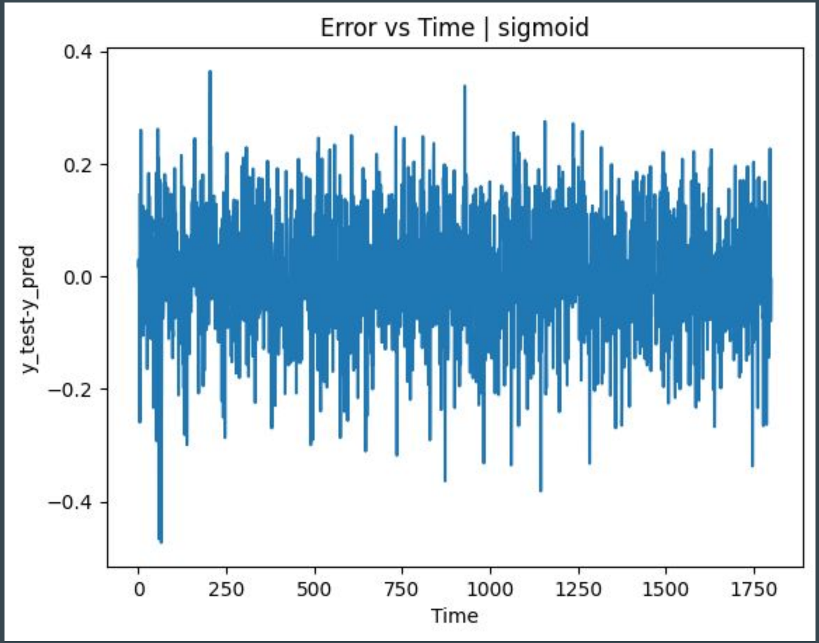
AIC for Linear Regression: -7305.426510529513

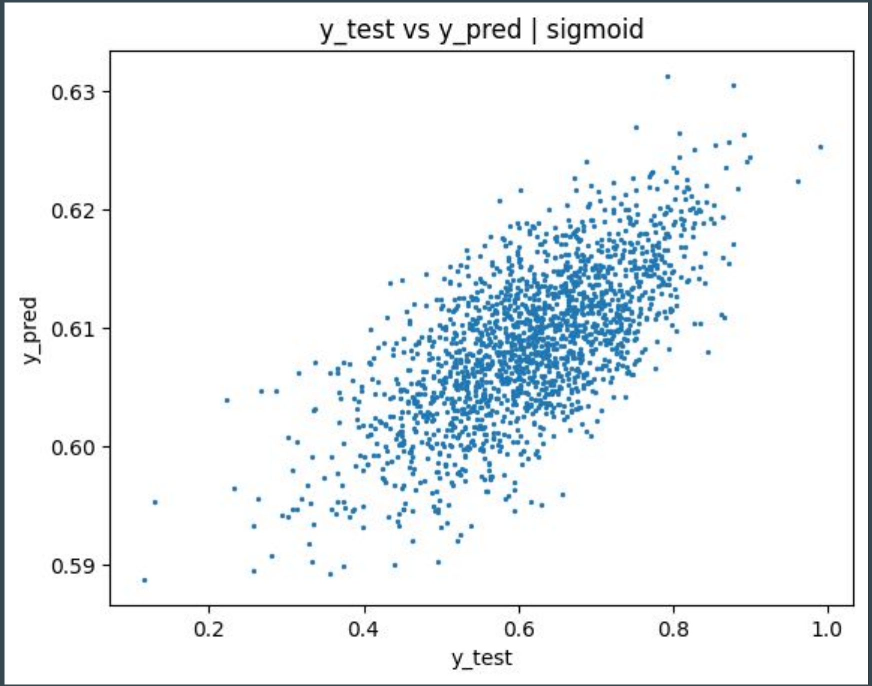
BIC for Linear Regression: 2811.866208161404





```
Activation Function:  sigmoid
Validation r2 score: 0.99685
Total no. of parameters: 1841
MSE: 0.011853492949070597
AIC for Linear Regression: -4301.238844161103
BIC for Linear Regression: 5816.053874529814
```





Nonlinear: Feed Forward (Deep Neural Network) | Effect of Number of Layers

This is the same as the Shallow counterparts, other than the fact that DNN has multiple hidden layers instead of just 1.

In order to isolate effect of number of layers we take number of nodes to be same in each hidden layer.

```
Layer Size:   15
```

Number of Layers = 30

```
Activation Function:  tanh
Validation r2 score: 0.15807
Total no. of parameters: 6796
MSE: 0.002211925662781832
AIC for Linear Regression: 2586.994755996653
BIC for Linear Regression: 39934.69780663406
```

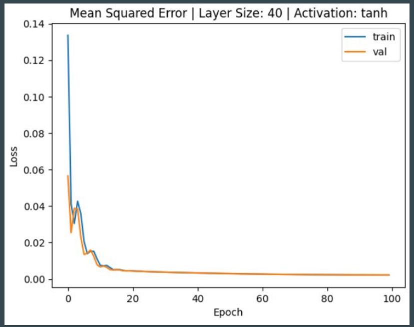
Number of Layers = 150

```
Activation Function:  tanh
Validation r2 score: 1.0
Total no. of parameters: 35596
MSE: 0.012794974889229245
AIC for Linear Regression: 63346.33501054779
BIC for Linear Regression: 258965.64604505178
```

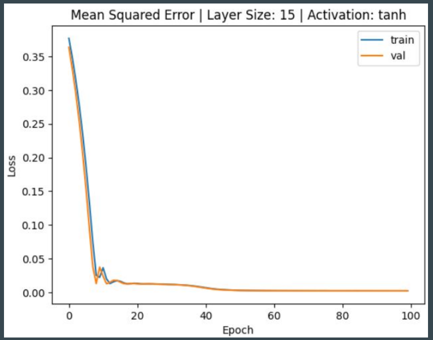
Number of Layers = 350

```
Activation Function:  tanh
Validation r2 score: 1.0
Total no. of parameters: 83596
MSE: 0.13438612739476824
AIC for Linear Regression: 163579.331464983
BIC for Linear Regression: 622984.6558059313
```

- Performance deteriorates for the same amount of training if bulkier models are used for simple problems such as this.
- The learning curves for models of similar size can be very different based on number of layers.



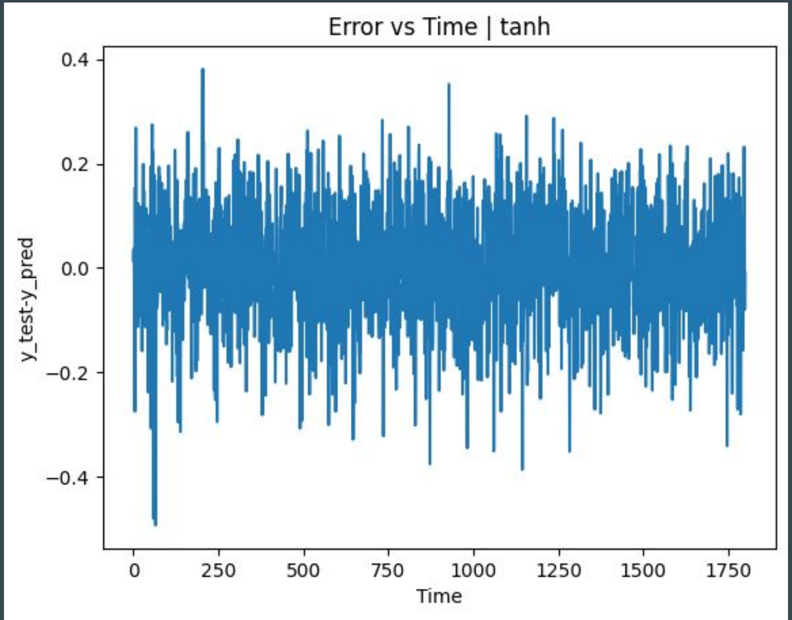
1 Hidden Layer, Parameters: 1841

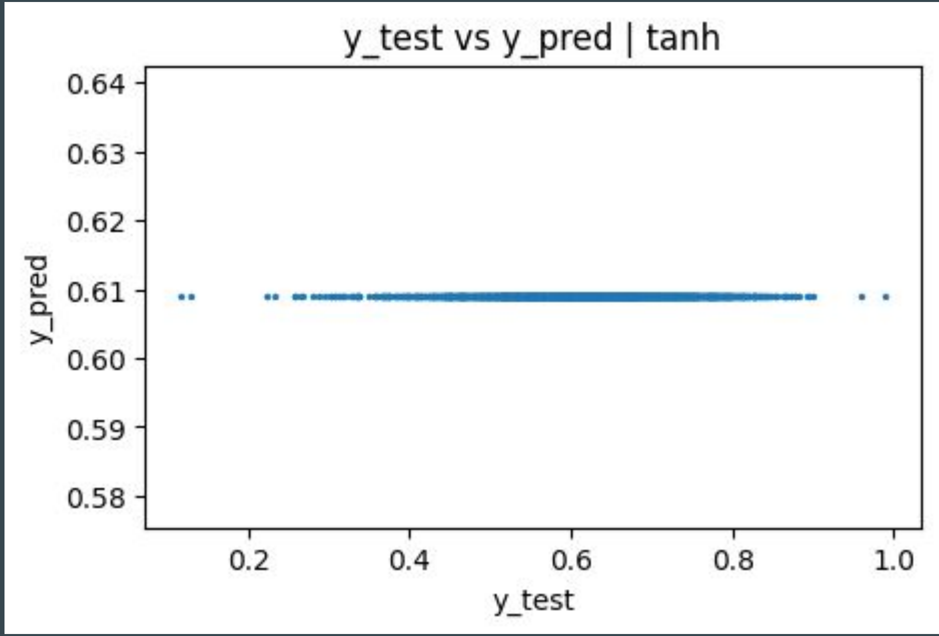


30 Hidden Layers, Parameters: 6796

Number of Layers: 150

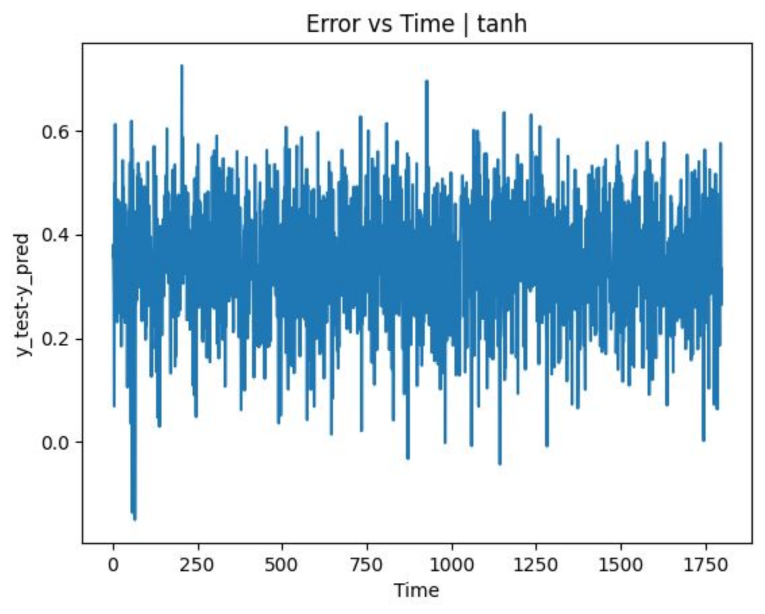
Activation Function: tanh
Validation r2 score: 1.0
Total no. of parameters: 35596
MSE: 0.012794974889229245
AIC for Linear Regression: 63346.33501054779
BIC for Linear Regression: 258965.64604505178

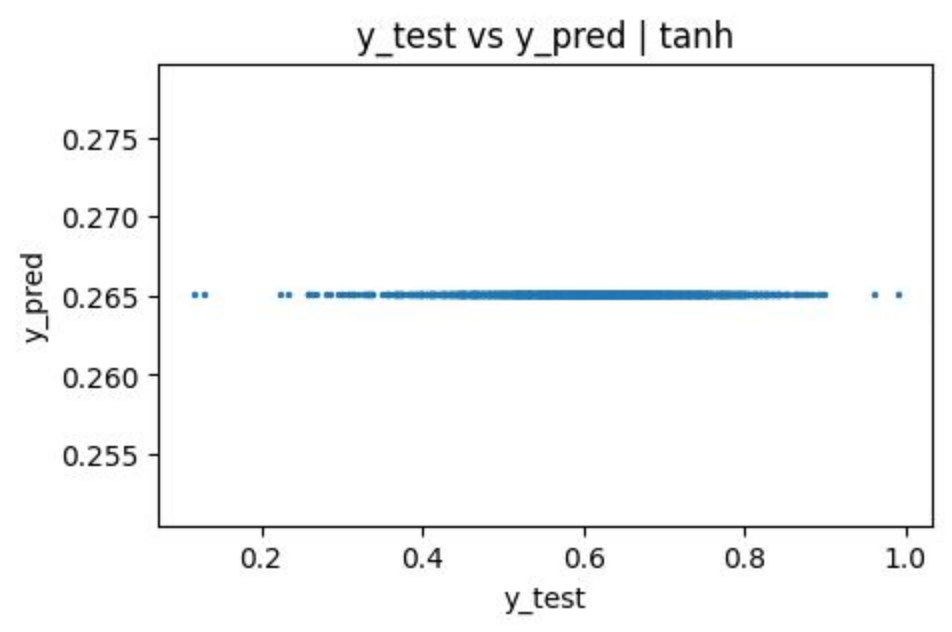




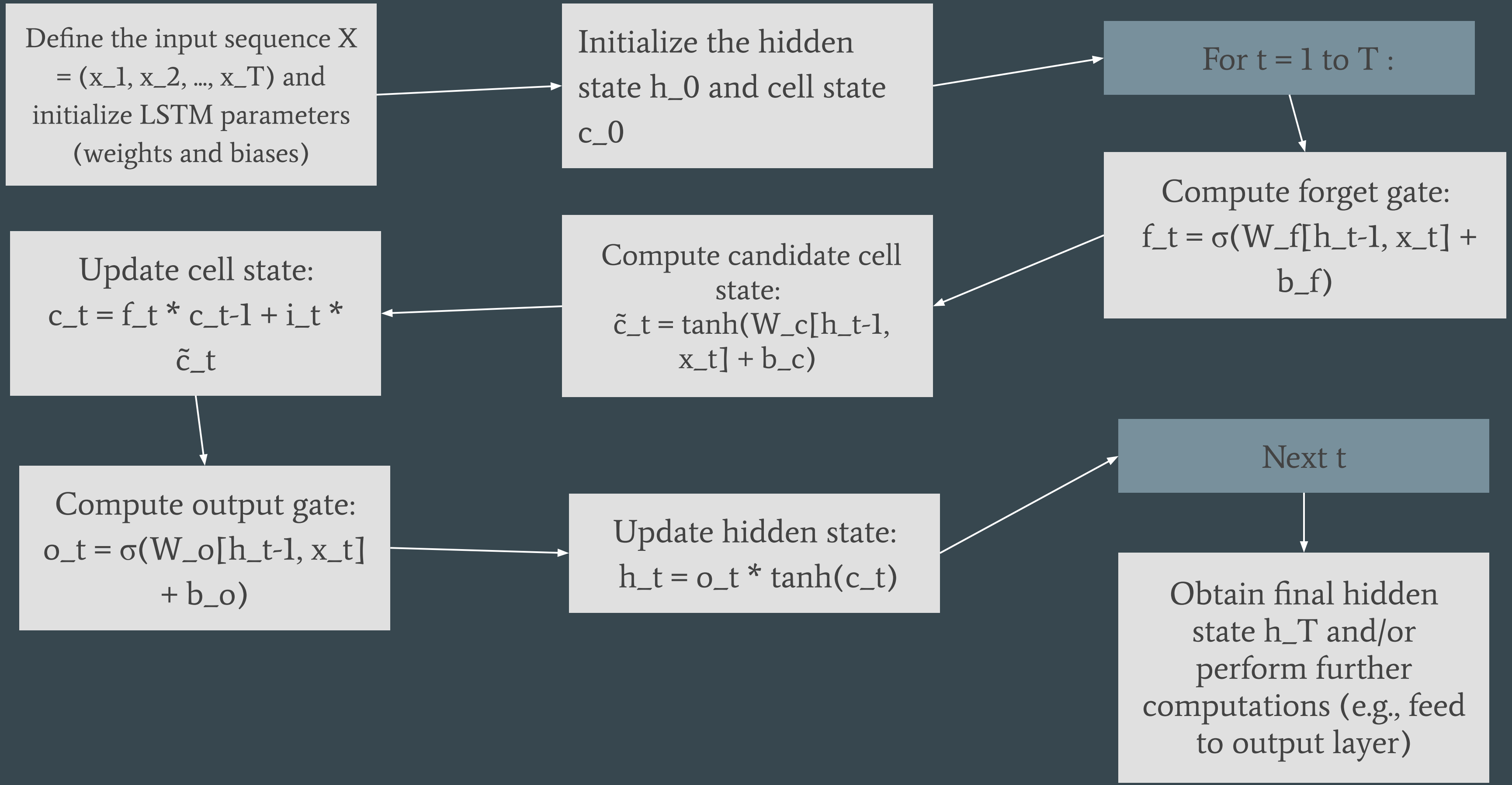
Number of Layers: 350

Activation Function: tanh
Validation r2 score: 1.0
Total no. of parameters: 83596
MSE: 0.13438612739476824
AIC for Linear Regression: 163579.331464983
BIC for Linear Regression: 622984.6558059313





Nonlinear: LSTM model



Nonlinear: LSTM model

Sequence Length: 18

```
Activation Function:  tanh
Validation r2 score: 0.98827
Total no. of parameters: 71851
MSE: 0.00014362540127040087
AIC for Linear Regression: 127775.05634973764
BIC for Linear Regression: 522635.2405597653
```

Sequence Length: 32

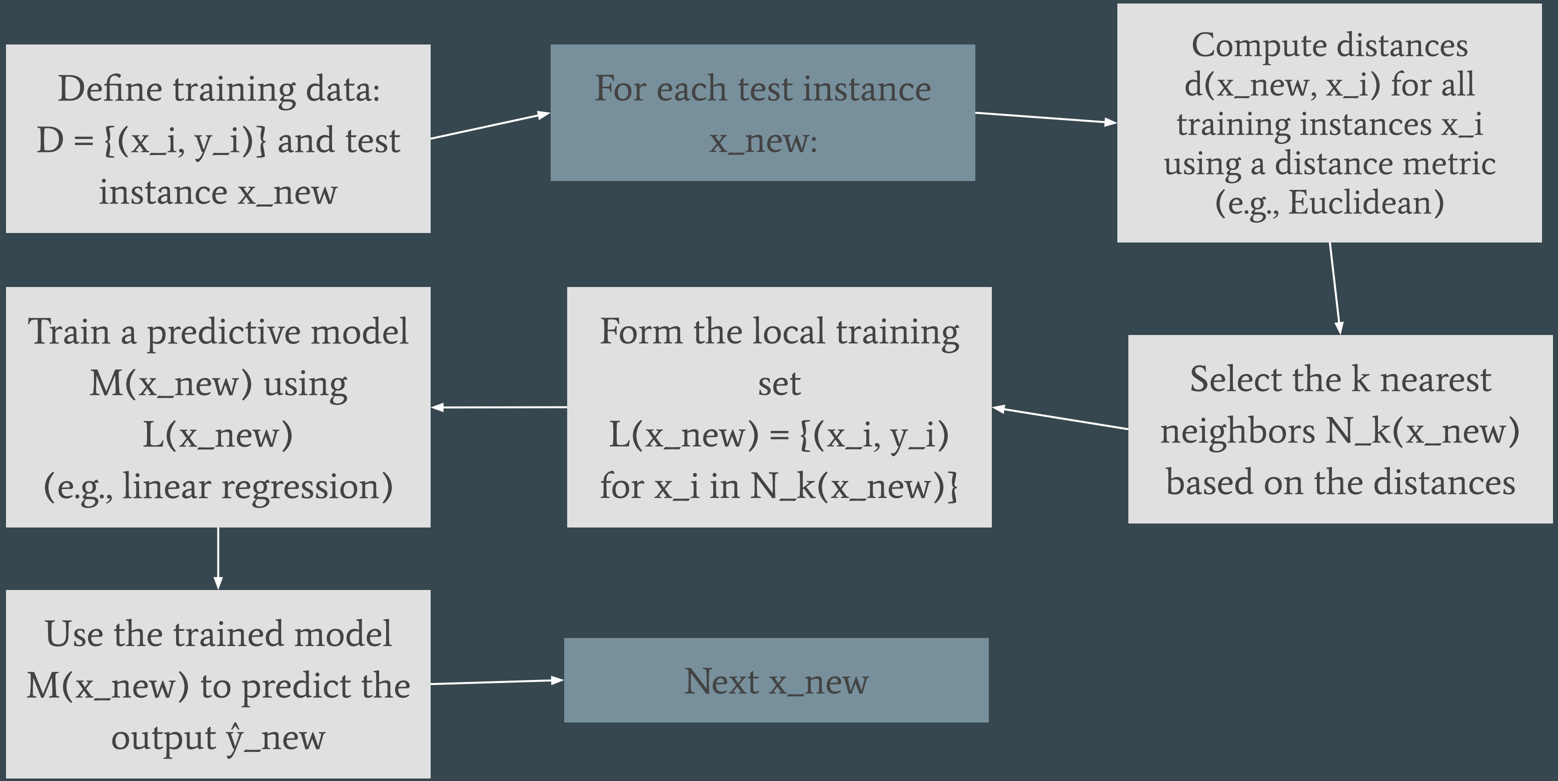
```
Activation Function:  tanh
Validation r2 score: 0.99273
Total no. of parameters: 71851
MSE: 0.00014879556106441684
AIC for Linear Regression: 127838.71291838013
BIC for Linear Regression: 522698.89712840784
```

Sequence Length: 56

```
Activation Function:  tanh
Validation r2 score: 0.99379
Total no. of parameters: 71851
MSE: 0.00018629797214371374
AIC for Linear Regression: 128243.30630243689
BIC for Linear Regression: 523103.49051246454
```

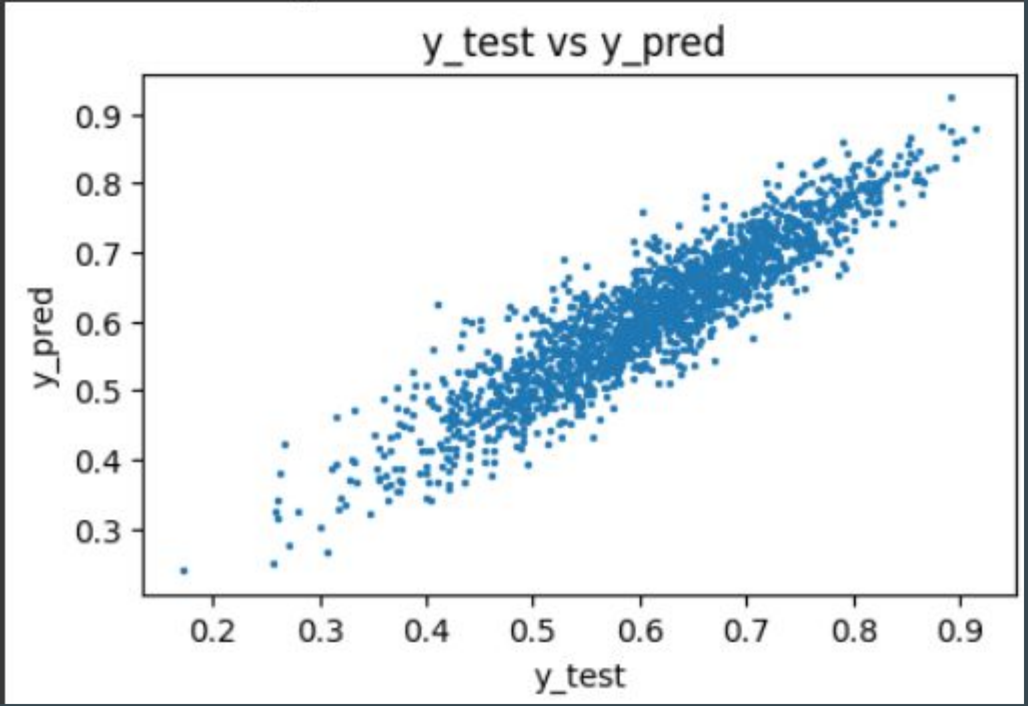
- The LSTM model shows significant improvement over DNN.
- The Sequence length was decided based on Auto-correlation of Target variable, which showed peaks on 17, 32 and 56.
- The accuracy improvement comes at significant computational cost.

Just-in-time learning based predictive model: using kNN



Just-in-time learning based predictive model: using kNN & Regression

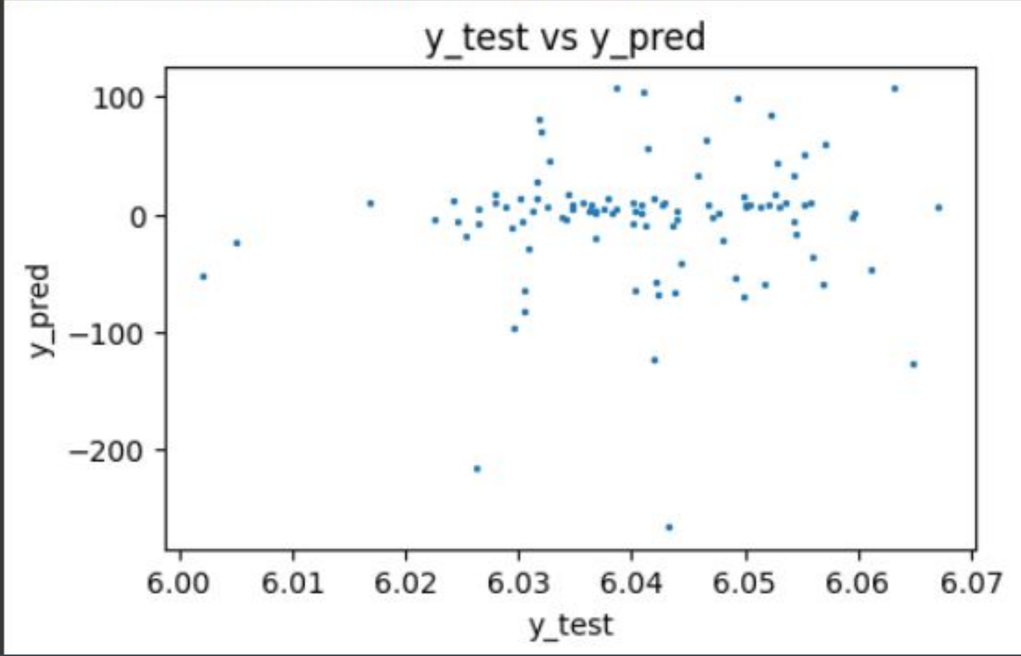
```
Total no. of parameters: 4
MSE: 0.0020526129881778274
AIC for Linear Regression: -11131.555004382732
BIC for Linear Regression: -11109.572836607194
```



- No significant improvement from normal regression since the data did not have multiple modes

Just-in-time learning based predictive model: using kNN & ANN

```
Activation Function:  tanh
Validation r2 score: -21007205.14218
Total no. of parameters: 316
MSE: 3060.6601742965117
AIC for Linear Regression: 1434.6385914911355
BIC for Linear Regression: 2257.8723702633724
```



We did not get a satisfactory output from KNN based JITLM.

However it can be due to the following reasons:

- Coding errors
- Unoptimised KNN - identification is slow.
- Tested on only 100 data points due to computation constraints.
- Not enough iterations in training

Conclusion:

What we inferred from this Project

Conclusion

- We trained and developed multiple models to optimize the monomer concentration for the given PMMA data set.

- We learnt the practical implementation of all the ML concepts taught in class.

- Various regression techniques showed us that on average, the models give more consistent and better results for linear regression methods, specifically subset selection.

- The model performed poorly if the penalty term was introduced (Lasso Regression).

- PCR performs worse than traditional regression for 2 components but has a slightly better performance for 4 or 5 components.

- LSTM provides the best model in terms of MSE but has high computational cost, so in terms of scalability, linear regression still provides the best model.

Acknowledgement

We would like to thank our professor, Jayaram Valluru sir for guiding us every step of the way and introducing us to the basics of Machine Learning. Its implementation in such a conclusive way would not have been possible without his invaluable guidance and support.

References

1) Comparison of unconstrained nonlinear state estimation techniques on a MMA polymer reactor ? Arjun V Shenoy, Vinay Prasad, Sirish L Shah (PMMA_Arjun)

2) Chen, T., Morris, J., and Martin, E. (2005). Particle filters for state and parameter estimation in batch processes.

3) Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models: Bikmukhametov, Jäschke