**Homework 7**

**Machine Learning I: Supervised Methods**

**EE 559**

Shivans Amattya

5982564472

A)

**Batch size = 32**

| Trail | Time (in sec. ) |
| --- | --- |
| 1 | 31 |
| 2 | 31 |
| 3 | 30 |
| 4 | 31 |
| 4 | 31 |

**Batch size = 64**

| Trail | Time (in sec. ) |
| --- | --- |
| 1 | 187 |
| 2 | 193 |
| 3 | 190 |
| 4 | 188 |
| 4 | 207 |

**Batch size = 512**

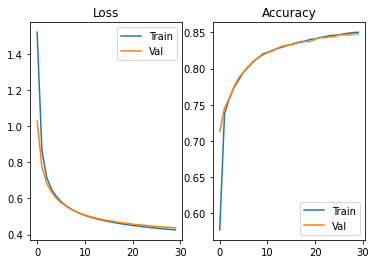
| Trail | Time (in sec. ) |
| --- | --- |
| 1 | 52 |
| 2 | 51 |
| 3 | 55 |
| 4 | 55 |
| 5 | 55 |

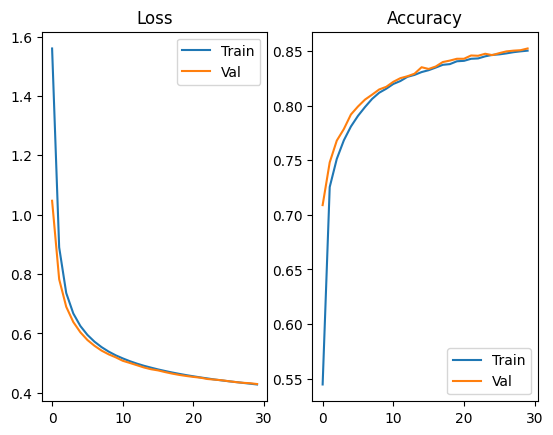
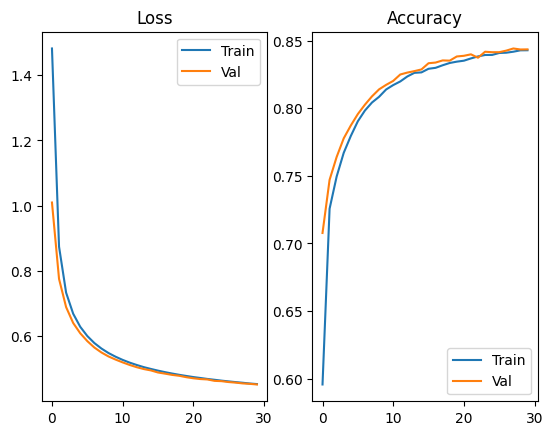
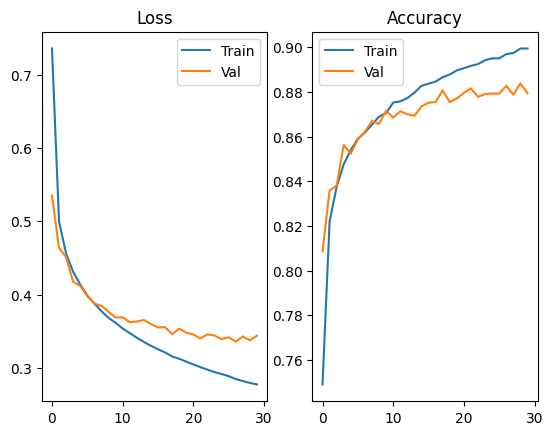
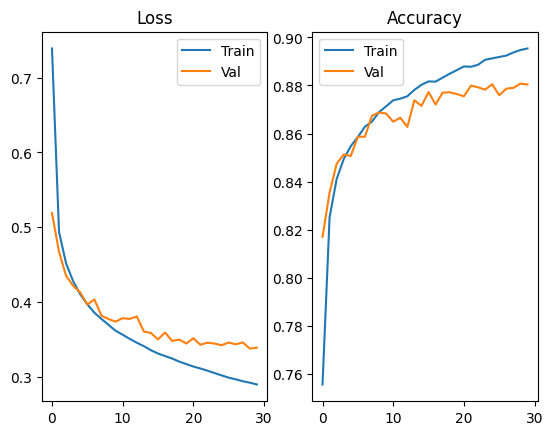
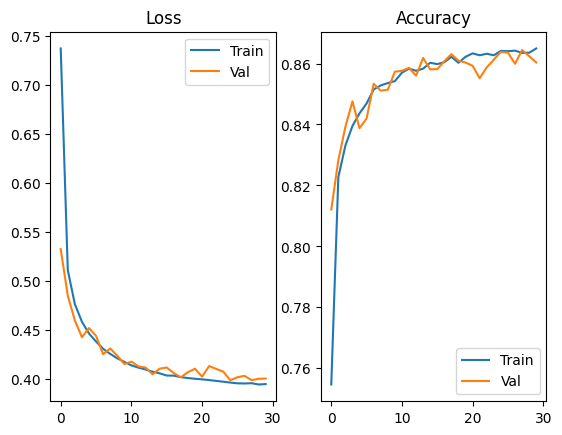
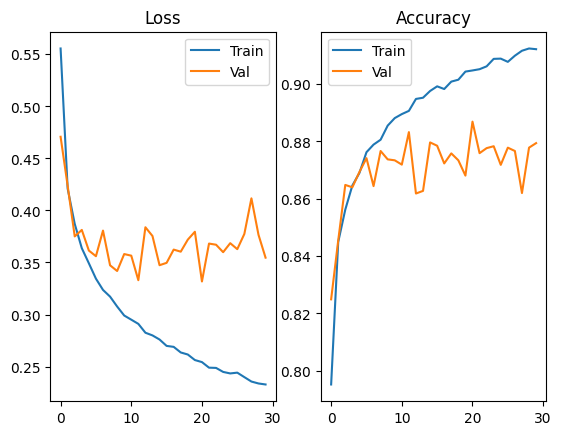
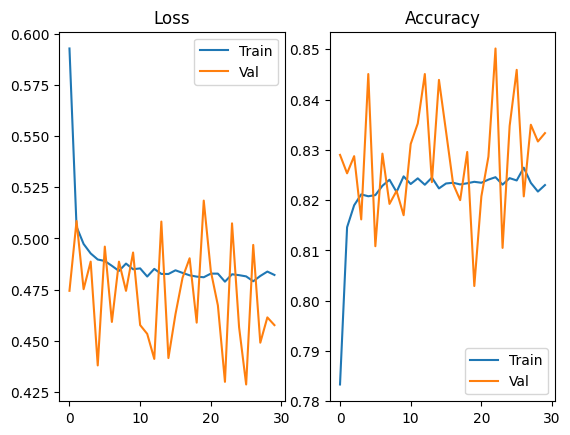
So the smallest batch size is 32 with the smallest runtime of 30 seconds

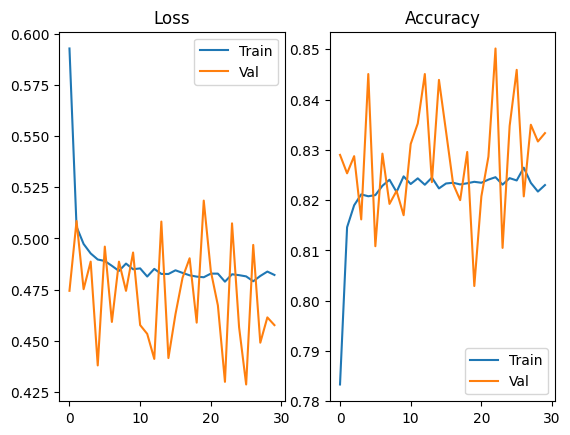
With **Mean** and **STD** as **30.8**, **0.39999999999999997** respectively.

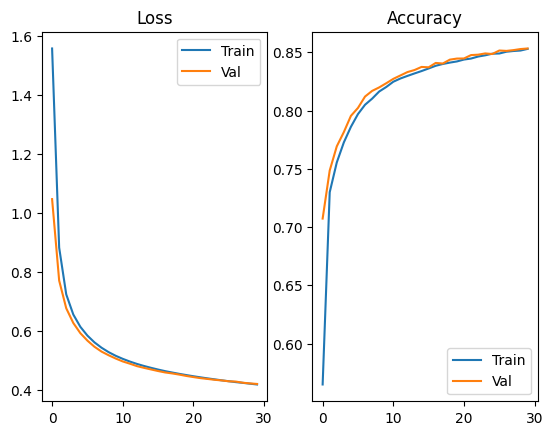
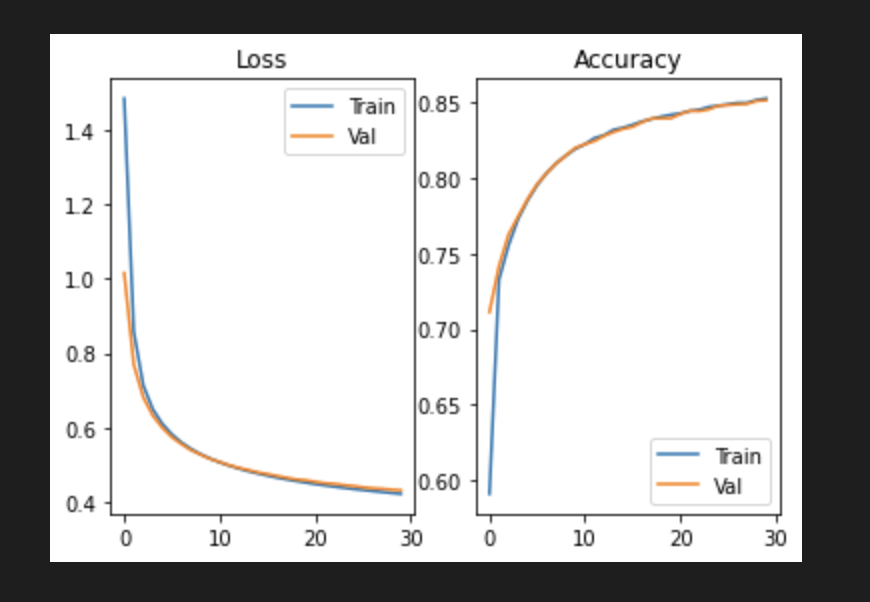
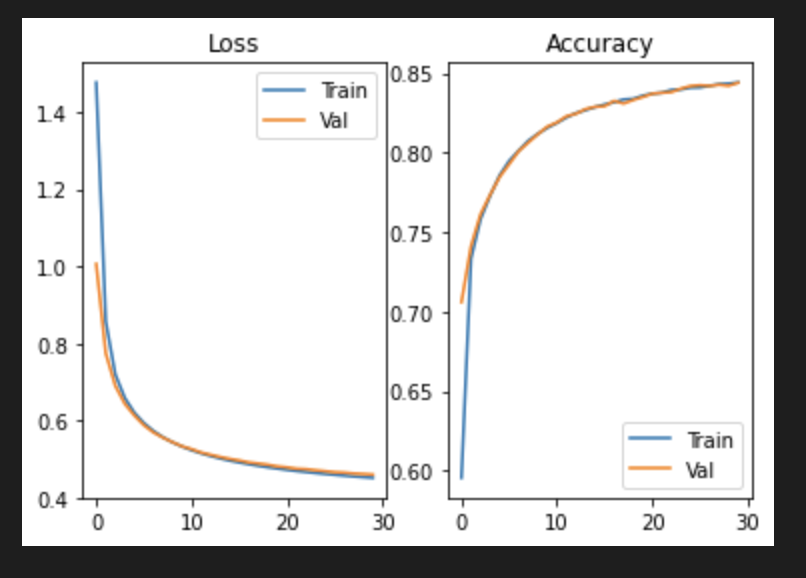
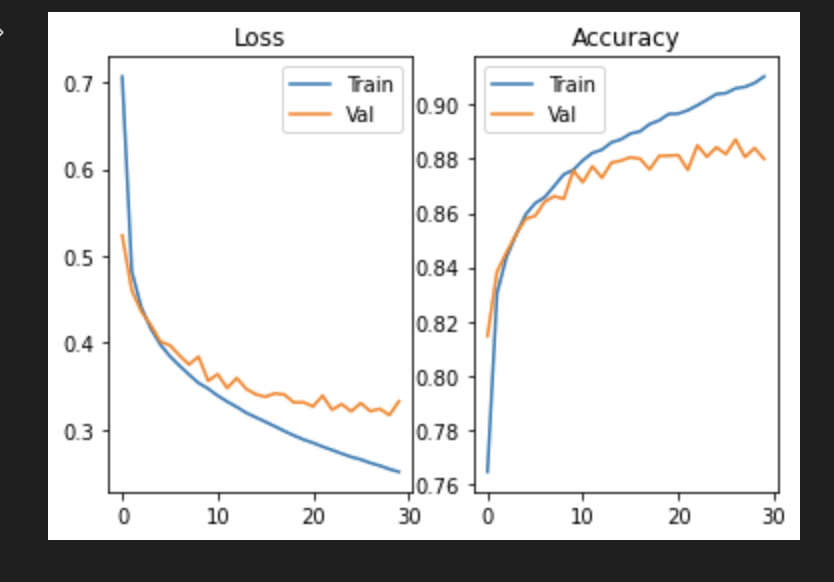
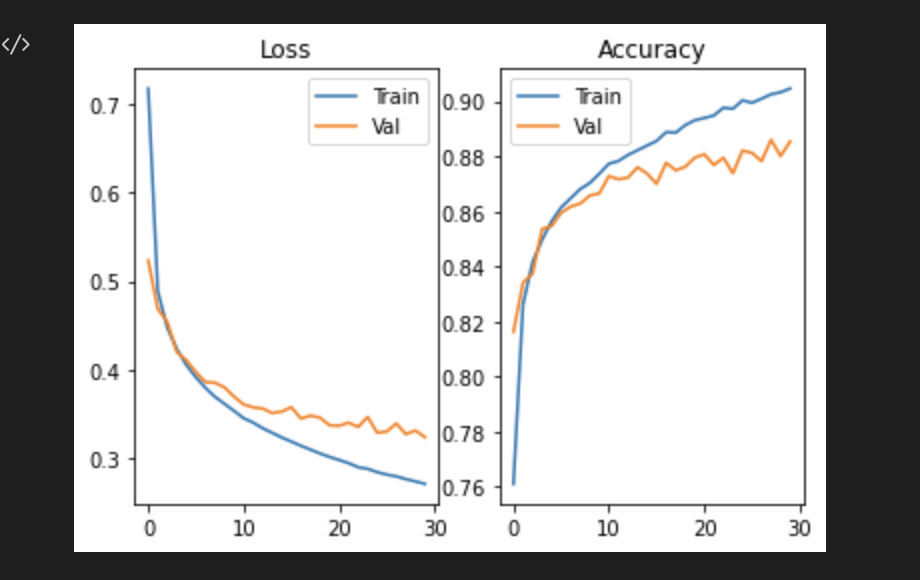
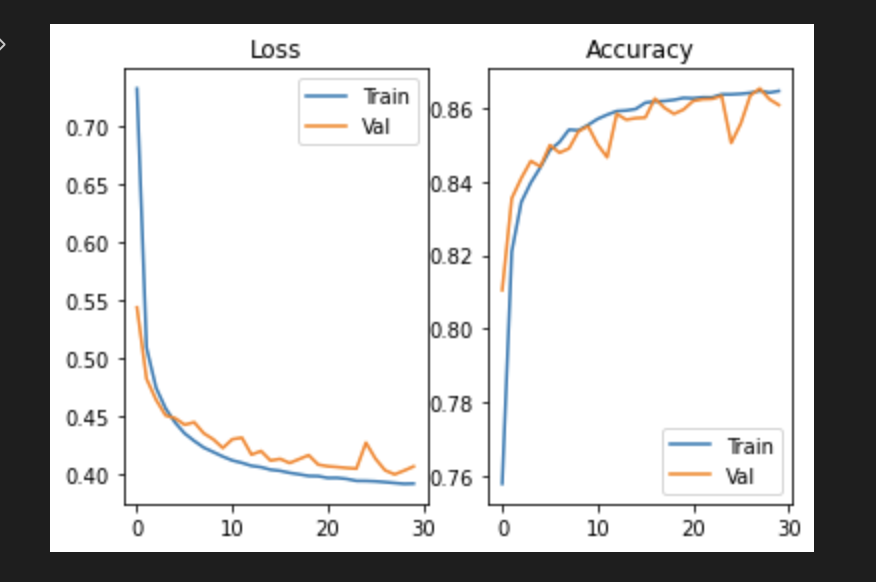
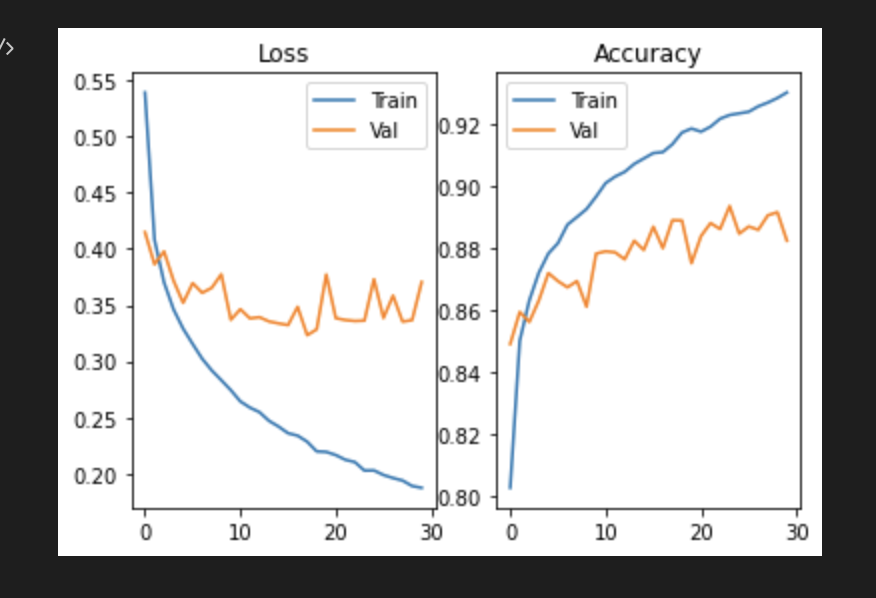
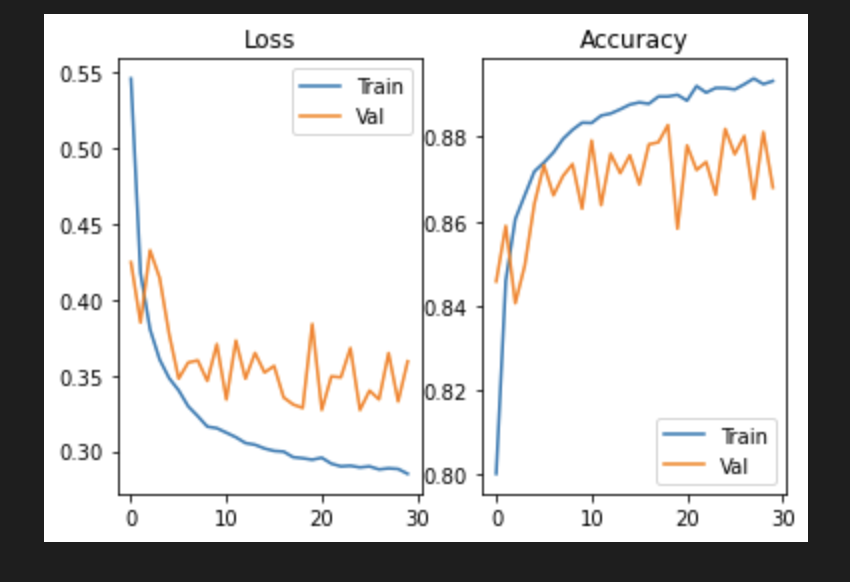
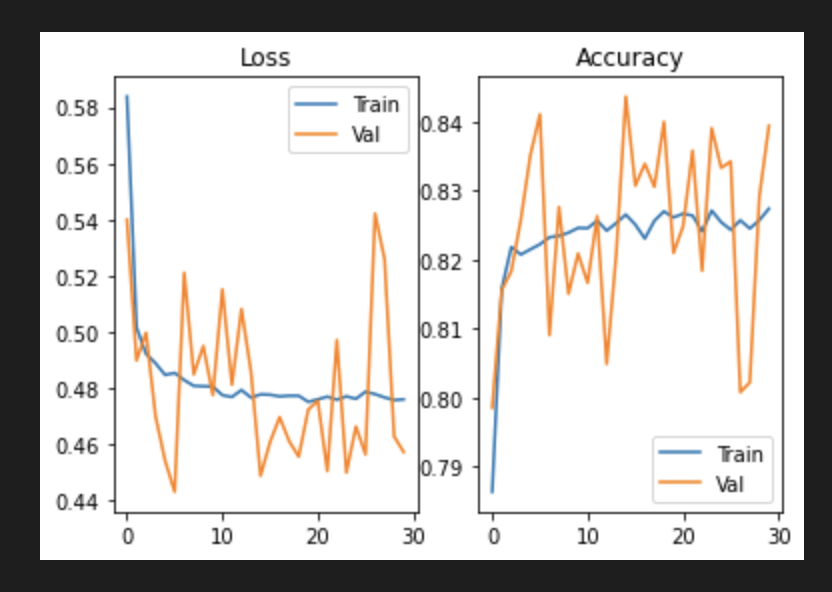
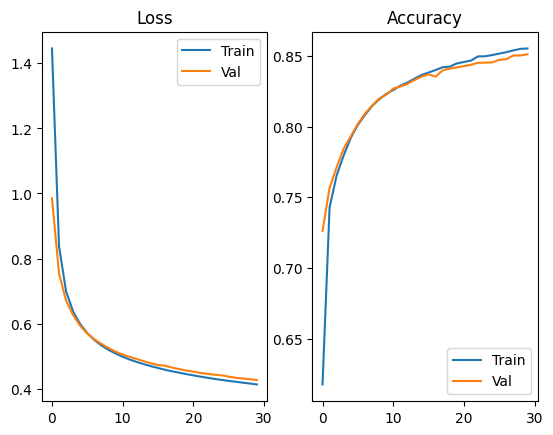
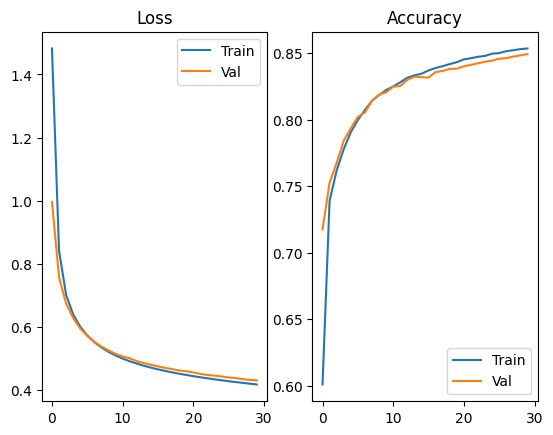
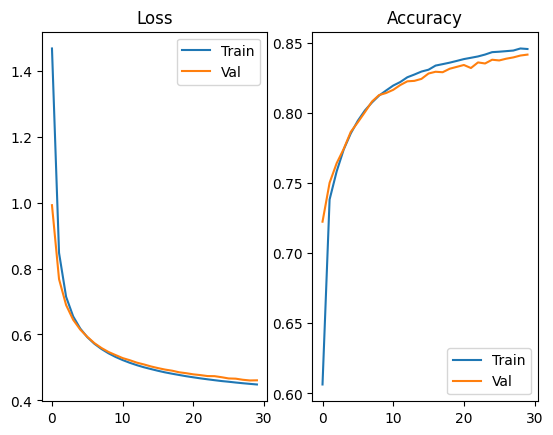
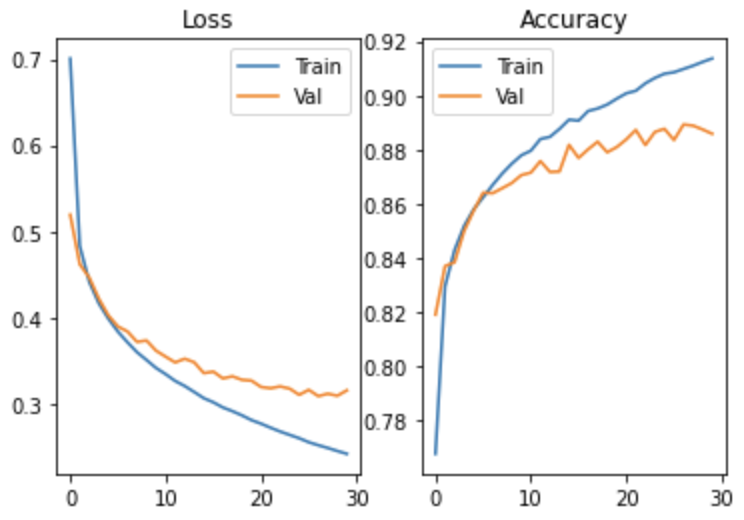
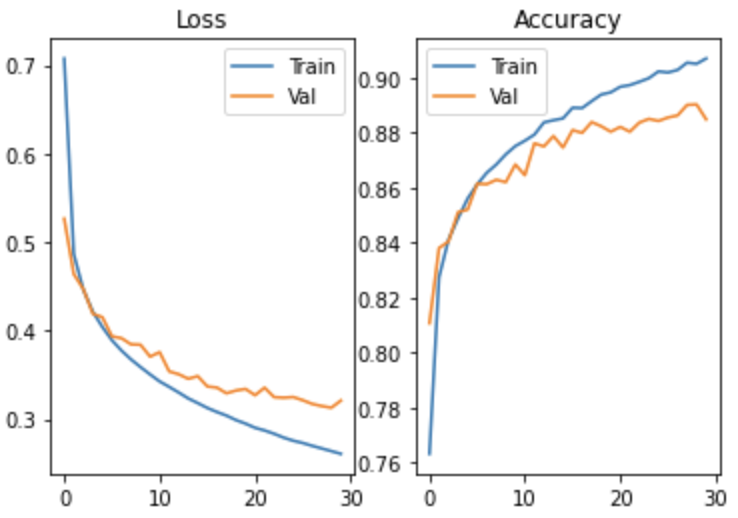
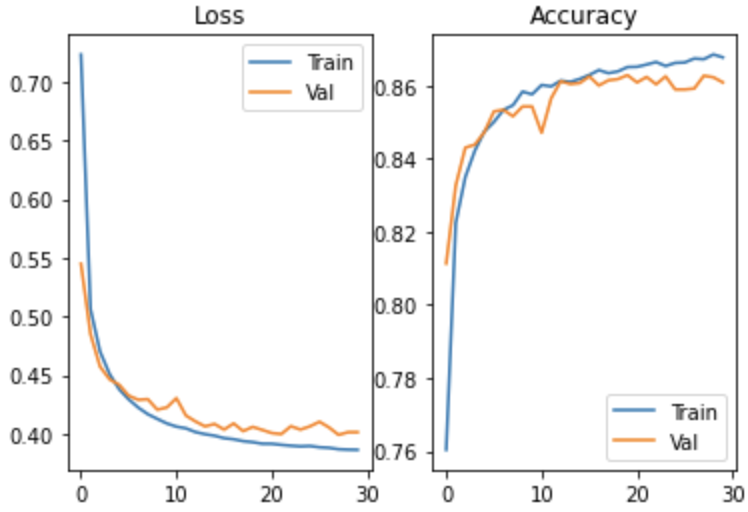
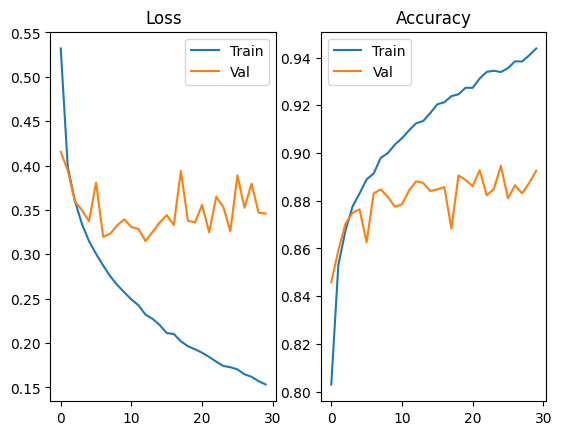
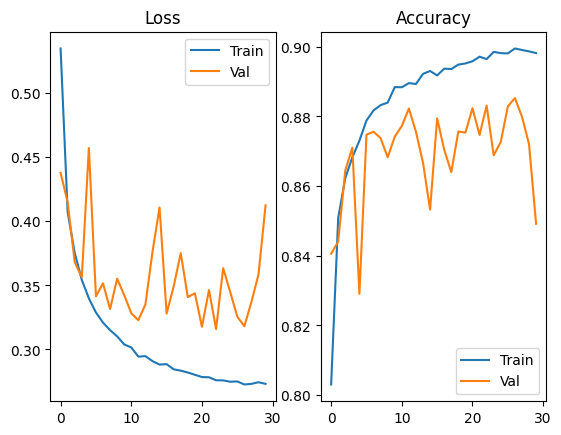
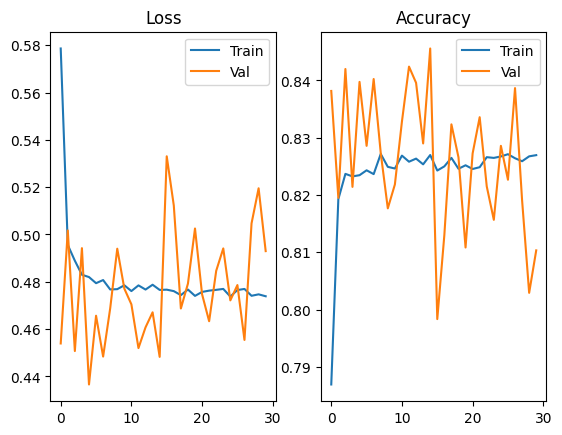
B)

* **M= 40 Learning Rate= 0.001 Weight Decay=0.0001**



* **M= 40 Learning Rate= 0.001 Weight Decay=0.001**
  + ****
* **M= 40 Learning Rate= 0.001 Weight Decay=0.01**
  + ****
* **M= 40 Learning Rate= 0.01 Weight Decay=0.0001**
  + ****
* **M= 40 Learning Rate= 0.01 Weight Decay=0.001**
  + ****
* **M= 40 Learning Rate= 0.01 Weight Decay=0.01**
  + ****
* **M= 40 Learning Rate= 0.1 Weight Decay=0.0001**
  + ****
* **M= 40 Learning Rate= 0.1 Weight Decay=0.001**
  + ****
* **M= 40 Learning Rate= 0.1 Weight Decay=0.01**

****

* **M= 80 Learning Rate= 0.001 Weight Decay=0.0001**
  + ****
* **M= 80 Learning Rate= 0.001 Weight Decay=0.001**
  + ****
* **M= 80 Learning Rate= 0.001 Weight Decay=0.01**
  + ****
* **M= 80 Learning Rate= 0.01 Weight Decay=0.0001**
  + ****
* **M= 80 Learning Rate= 0.01 Weight Decay=0.001**
  + ****
* **M= 80 Learning Rate= 0.01 Weight Decay=0.01**
  + ****
* **M= 80 Learning Rate= 0.1 Weight Decay=0.0001**
  + ****
* **M= 80 Learning Rate= 0.1 Weight Decay=0.001**
  + ****
* **M= 80 Learning Rate= 0.1 Weight Decay=0.01**
  + ****
* **M= 160 Learning Rate= 0.001 Weight Decay=0.0001**
  + ****
* **M= 160 Learning Rate= 0.001 Weight Decay=0.001**
  + ****
* **M= 160 Learning Rate= 0.001 Weight Decay=0.01**
  + ****
* **M= 160 Learning Rate= 0.01 Weight Decay=0.0001**
  + ****
* **M= 160 Learning Rate= 0.01 Weight Decay=0.001**
  + ****
* **M= 160 Learning Rate= 0.01 Weight Decay=0.01**
  + ****
* **M= 160 Learning Rate= 0.1 Weight Decay=0.0001**
  + ****
* **M= 160 Learning Rate= 0.1 Weight Decay=0.001**
  + ****
* **M= 160 Learning Rate= 0.1 Weight Decay=0.01**
  + ****

C)For the best hyper-parameters found in part B is **M=160 lr=0.1 wd=0.0001** with accuracies[0.8949166666666667,0.8959166666666667,0.8966666666666666,0.8956666666666667,0.8954166666666666.

* **Mean** = 0.8957166666666666
* **Std** = 0.0005787918451394897
* **Max Accuracy** = 0.8966666666666666

D)

**Test accuracy**: 0.8880

----------------------------------------------------------------

Layer (type) Output Shape Param #

================================================================

Linear-1 [-1, 160]  **125,600**

ReLU-2 [-1, 160] 0

Linear-3 [-1, 10] 1,610

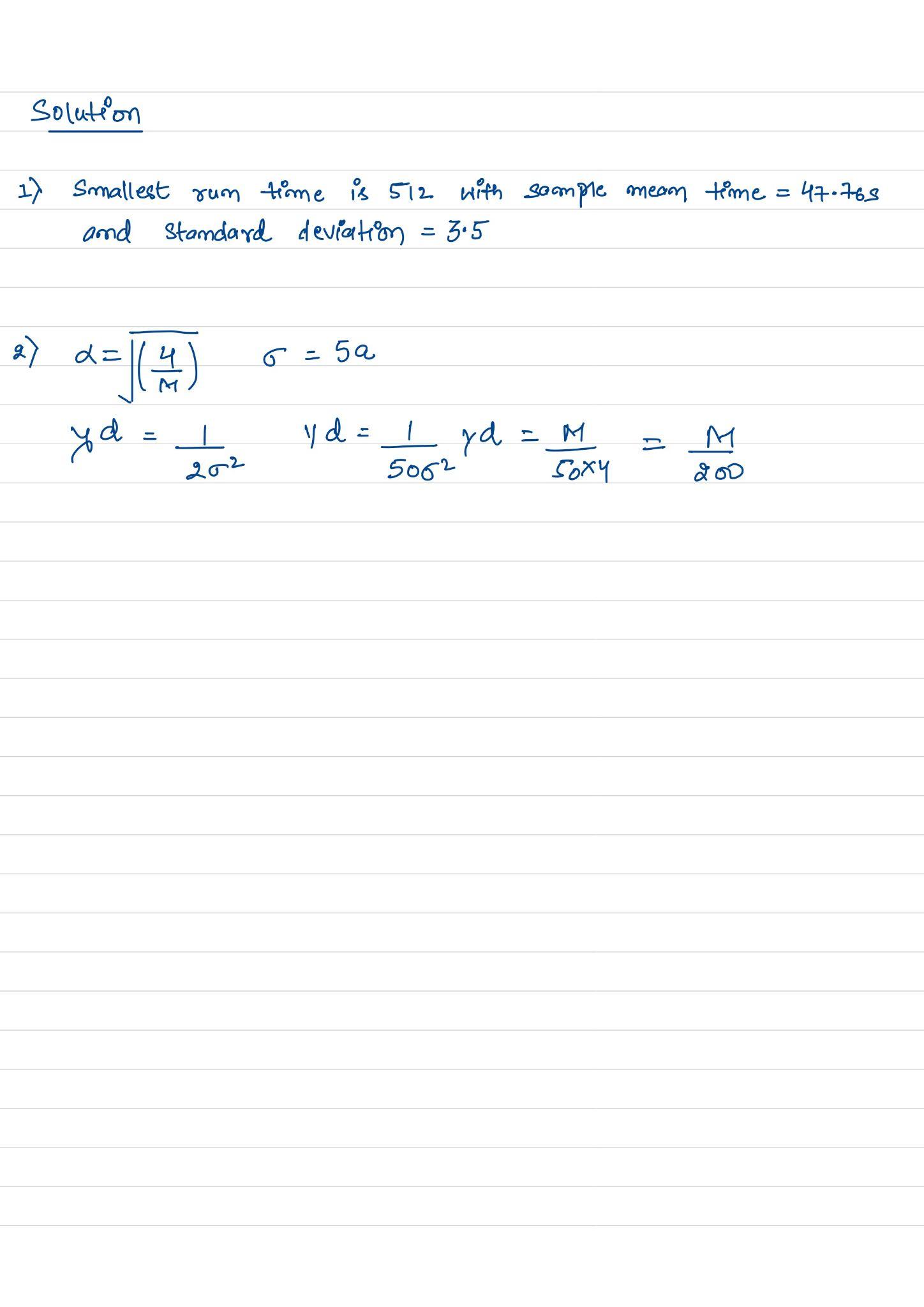
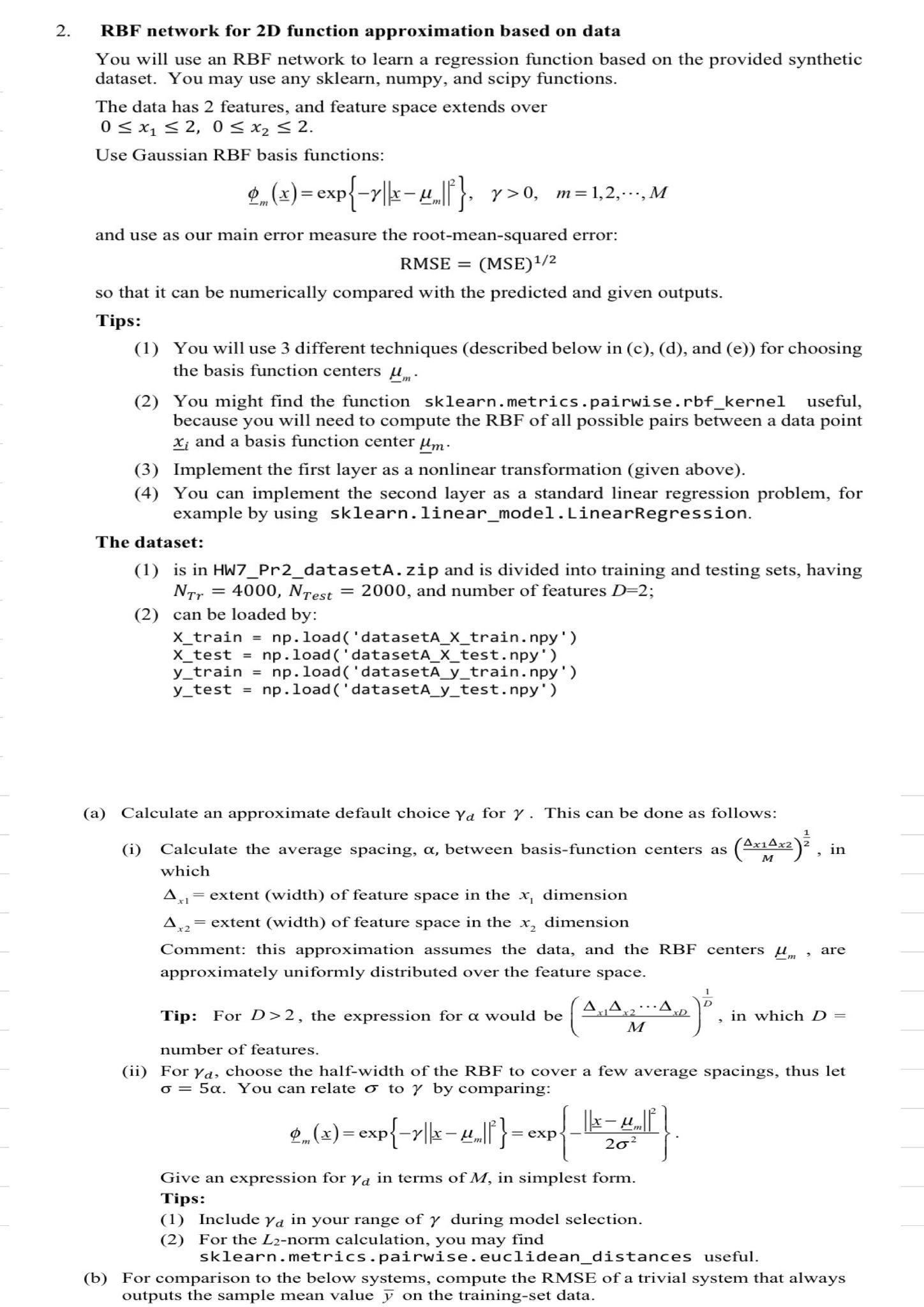
================================================================

**Total params**: 127,210

**Trainable params**: 127,210

Non-trainable params: 0

----------------------------------------------------------------



Solution 2.

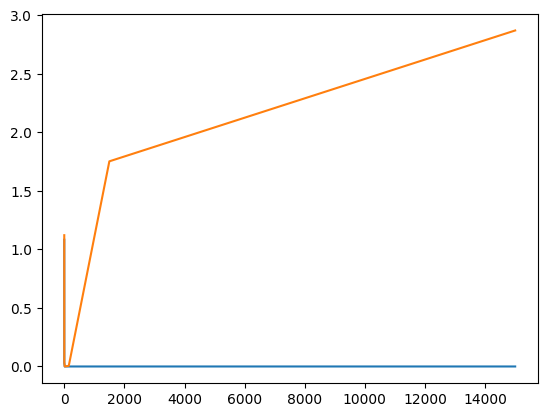
B) Rmse is 3.2035150890062902

C) Best value of Gamma = 15

| Gamma | Mean Rmse |
| --- | --- |
| 0.15 | 1.12084791e+00 |
| 1.5 | 3.48508178e-02 |
| 15 | 8.14464632e-07 |
| 150 | 5.33717408e-03 |
| 1500 | 1.75161082e+00 |
| 15000 | 2.86882164e+00 |

* Best Mean is 8.14464632e-07 at Gamma = 15

| * Gamma | * STD of RMSE |
| --- | --- |
| * 0.15 | * 2.76836370e-02 |
| * 1.5 | * 3.37456989e-03 |
| * 15 | * 7.43958219e-07 |
| * 150 | * 3.28757045e-03 |
| * 1500 | * 3.50427056e-01 |
| * 15000 | * 4.76147431e-02 |



* **RMSE VS Gammma**

D) At gamma =30 and M=600, we get the best values that is 3.46000610e-03

**For M=30**

| Gamma | M | Mean of Rmse |
| --- | --- | --- |
| 0.0015 | 30 | 4.35742330e+00 |
| 0.015 | 30 | 2.06722118e+00 |
| 0.15 | 30 | .58203626e+00 |
| 1.5 | 30 | 1.58787466e+00 |
| 15 | 30 | 1.69986679e+00 |
| 150 | 30 | 2.83534066e+00 |

**For M= 60**

| Gamma | M | Mean of RMSE |
| --- | --- | --- |
| 0.003 | 60 | 2.16869212e+00 |
| 0.03 | 60 | 1.58071675e+00 |
| 0.3 | 60 | 1.16296875e+00 |
| 3 | 60 | 8.14434687e-01 |
| 30 | 60 | 1.38844185e+00 |
| 300 | 60 | 2.77549267e+00 |

**For M = 100**

| Gamma | M | Mean of RMSE |
| --- | --- | --- |
| 0.0005 | 100 | 3.20880511e+00 |
| 0.005 | 100 | 1.44946814e+00 |
| 0.5 | 100 | 5.80309893e-01 |
| 5 | 100 | 1.85780127e-01 |
| 50 | 100 | 9.96879986e-01 |
| 500 | 100 | 2.80077308e+00 |

**For M = 300**

| Gamma | M | Mean of RMSE |
| --- | --- | --- |
| 0.015 | 300 | 2.49712764e+00 |
| 0.15 | 300 | 1.21113499e+00 |
| 1.5 | 300 | 4.65652109e-02 |
| 15 | 300 | 8.55723275e-03 |
| 150 | 300 | 1.09787050e+00 |
| 1500 | 300 | 2.79213959e+00 |

**For M = 600**

| Gamma | M | Mean of RMSE |
| --- | --- | --- |
| 0.03 | 600 | 1.82323381e+00 |
| 0.3 | 600 | 7.09406792e-01 |
| 3 | 600 | 3.58462533e-03 |
| 30 | 600 | 3.46000610e-03 |
| 300 | 600 | 1.08293355e+00 |
| 3000 | 600 | 2.81337570e+00 |

**Standard Deviation of Rmse**

**For M=30**

| Gamma | M | STD of Rmse |
| --- | --- | --- |
| 0.0015 | 30 | 1.59293532e+00 |
| 0.015 | 30 | 1.06052451e-01 |
| 0.15 | 30 | 5.29621517e-02 |
| 1.5 | 30 | 7.15057785e-02 |
| 15 | 30 | 2.34460748e-01 |
| 150 | 30 | 4.30337911e-02 |

**For M= 60**

| Gamma | M | STD of RMSE |
| --- | --- | --- |
| 0.003 | 60 | 8.56905833e-02 |
| 0.03 | 60 | 1.13569065e-01 |
| 0.3 | 60 | 1.29663211e-02 |
| 3 | 60 | 5.75970295e-02 |
| 30 | 60 | 3.14965699e-01 |
| 300 | 60 | 1.19676132e-01 |

**For M = 100**

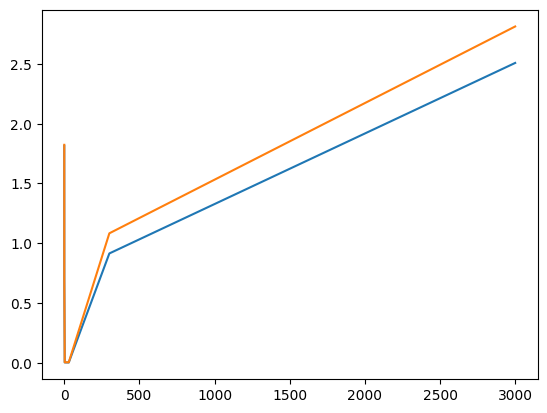
| Gamma | M | STD of RMSE |
| --- | --- | --- |
| 0.0005 | 100 | 4.18586209e-01 |
| 0.005 | 100 | 3.77875619e-02 |
| 0.5 | 100 | 1.89561229e-02 |
| 5 | 100 | 2.88074843e-02 |
| 50 | 100 | 1.32735705e-01 |
| 500 | 100 | 1.02756234e-01 |

**For M = 300**

| Gamma | M | STD of RMSE |
| --- | --- | --- |
| 0.015 | 300 | 7.40388745e-01 |
| 0.15 | 300 | 2.19778685e-02 |
| 1.5 | 300 | 2.79729179e-03 |
| 15 | 300 | 1.91300547e-03 |
| 150 | 300 | 1.91300547e-03 |
| 1500 | 300 | 6.65150525e-02 |

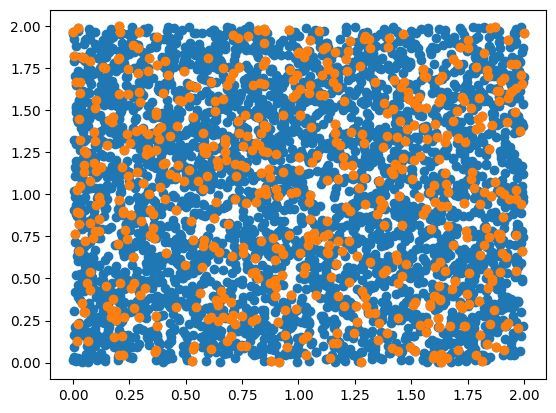
**For M = 600**

| Gamma | M | STD of RMSE |
| --- | --- | --- |
| 0.03 | 600 | 2.66149835e-01 |
| 0.3 | 600 | 2.50104216e-02 |
| 3 | 600 | 5.73573440e-04 |
| 30 | 600 | 5.95122357e-04 |
| 300 | 600 | 1.73065908e-02 |
| 3000 | 600 | 4.66475279e-02 |

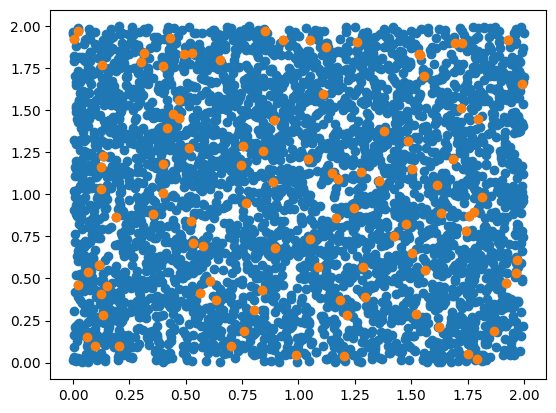


* **RMSE vs Gamma**
* If computational cost is a concern, the smallest value of M in RMSE is at least a factor of 10 lower than 100, and the Value of Gamma should be equal to 5. The number of hidden units would be reduced by a factor of 30.

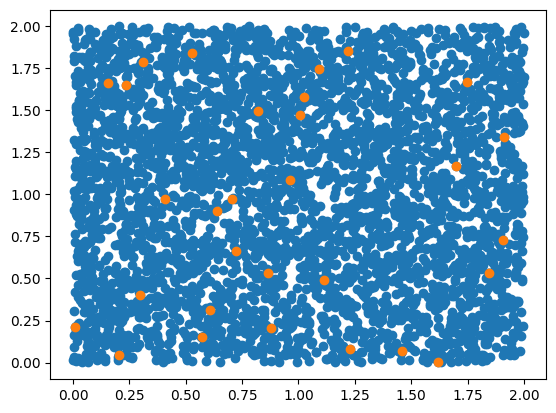
**Best Model**



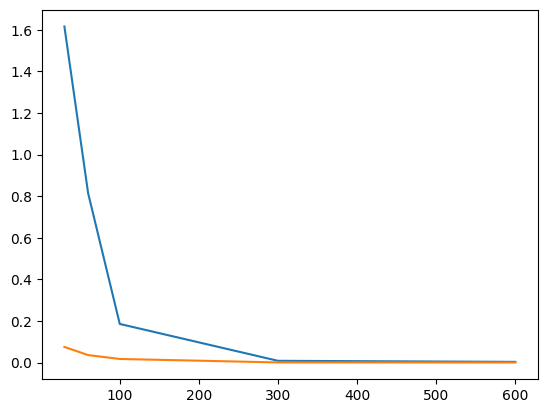
**RCC Model**



**Lowest Model**



**M vs. Rmse**



E) At gamma =30 and M=600, we get the best values that is 2.76910425e-03

**For M=30**

| Gamma | M | Mean of Rmse |
| --- | --- | --- |
| 0.0015 | 30 | 3.57778198e+00 |
| 0.015 | 30 | 2.37880375e+00 |
| 0.15 | 30 | 1.55541207e+00 |
| 1.5 | 30 | 1.64606473e+00 |
| 15 | 30 | 1.69528336e+00 |
| 150 | 30 | 2.86239474e+00 |

**For M= 60**

| Gamma | M | Mean of RMSE |
| --- | --- | --- |
| 0.003 | 60 | 2.22855485e+00 |
| 0.03 | 60 | 1.77715764e+00 |
| 0.3 | 60 | 1.17614559e+00 |
| 3 | 60 | 8.54447216e-01 |
| 30 | 60 | 7.59822003e-01 |
| 300 | 60 | 2.72635351e+00 |

**For M = 100**

| Gamma | M | Mean of RMSE |
| --- | --- | --- |
| 0.0005 | 100 | 3.03776548e+00 |
| 0.005 | 100 | 1.47399360e+00 |
| 0.5 | 100 | 5.76093283e-01 |
| 5 | 100 | 1.65324743e-01 |
| 50 | 100 | 5.30324799e-01 |
| 500 | 100 | 2.70095018e+00 |

**For M = 300**

| Gamma | M | Mean of RMSE |
| --- | --- | --- |
| 0.015 | 300 | 2.25797464e+00 |
| 0.15 | 300 | 1.20240097e+00 |
| 1.5 | 300 | 4.30371041e-02 |
| 15 | 300 | 5.25558291e-03 |
| 150 | 300 | 4.54740479e-01 |
| 1500 | 300 | 2.76662067e+00 |

**For M = 600**

| Gamma | M | Mean of RMSE |
| --- | --- | --- |
| 0.03 | 600 | 1.63347598e+00 |
| 0.3 | 600 | 7.08320641e-01 |
| 3 | 600 | 3.24497975e-03 |
| 30 | 600 | 2.76910425e-03 |
| 300 | 600 | 5.41288262e-01 |
| 3000 | 600 | 4.55111541e+00 |

**Standard Deviation of Rmse**

**For M=30**

| Gamma | M | STD of Rmse |
| --- | --- | --- |
| 0.0015 | 30 | 9.15161991e-01 |
| 0.015 | 30 | 3.01923276e-01 |
| 0.15 | 30 | 7.21690898e-02 |
| 1.5 | 30 | 2.51382581e-02 |
| 15 | 30 | 4.98761505e-02 |
| 150 | 30 | 3.00176785e-02 |

**For M= 60**

| Gamma | M | STD of RMSE |
| --- | --- | --- |
| 0.003 | 60 | 1.00716378e-01 |
| 0.03 | 60 | 1.54352793e-01 |
| 0.3 | 60 | 1.89714863e-02 |
| 3 | 60 | 1.31277669e-02 |
| 30 | 60 | 1.76810163e-01 |
| 300 | 60 | 6.28110206e-02 |

**For M = 100**

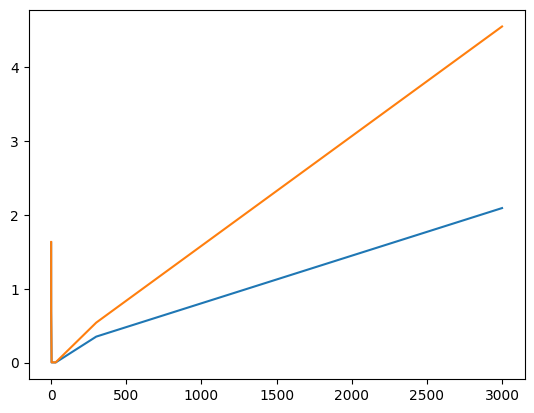
| Gamma | M | STD of RMSE |
| --- | --- | --- |
| 0.0005 | 100 | 5.18523854e-01 |
| 0.005 | 100 | 2.11635553e-02 |
| 0.5 | 100 | 2.47157915e-02 |
| 5 | 100 | 9.63951128e-03 |
| 50 | 100 | 6.82359605e-02 |
| 500 | 100 | 3.68659908e-02 |

**For M = 30**

| Gamma | M | STD of RMSE |
| --- | --- | --- |
| 0.015 | 300 | 3.40700053e-01 |
| 0.15 | 300 | 2.63073660e-02 |
| 1.5 | 300 | 2.23359602e-03 |
| 15 | 300 | 5.88138892e-04 |
| 150 | 300 | 3.98117004e-02 |
| 1500 | 300 | 4.57810724e-02 |

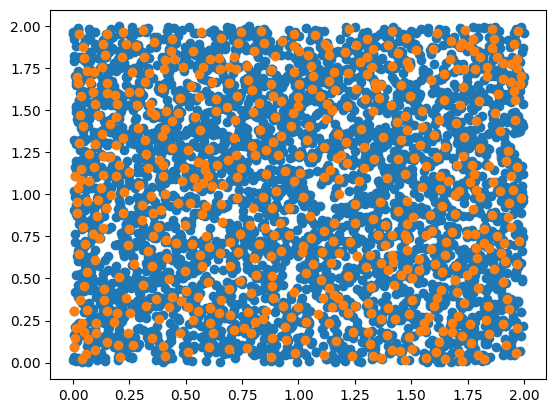
**For M = 600**

| Gamma | M | STD of RMSE |
| --- | --- | --- |
| 0.03 | 600 | 6.94368929e-02 |
| 0.3 | 600 | 2.55777318e-02 |
| 3 | 600 | 3.83608723e-04 |
| 30 | 600 | 5.57107987e-04 |
| 300 | 600 | 2.32873165e-02 |
| 3000 | 600 | 1.65863734e+00 |

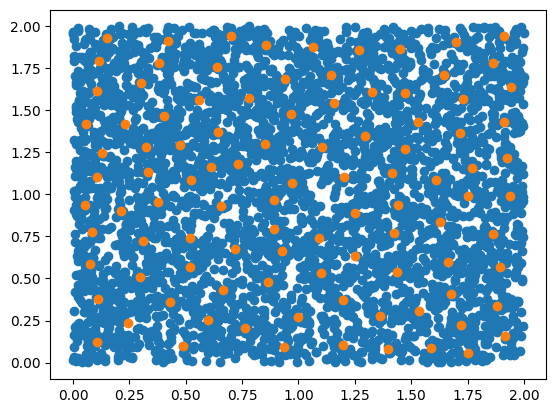


* **RMSE vs Gamma**
* If computational cost is a concern, the smallest value of K in RMSE is at least a factor of 10 lower than 100, and the Value of Gamma should be equal to 5. The number of hidden units would be reduced by a factor of 30.

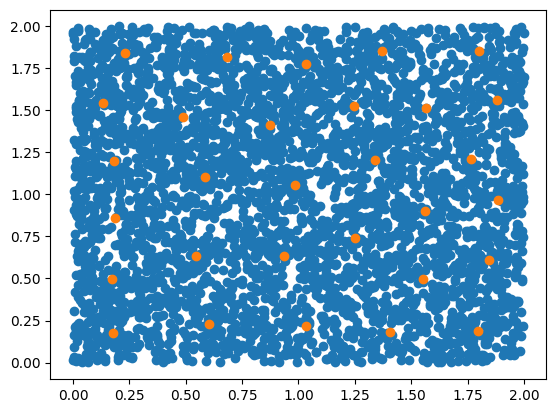
**Best Model**

****

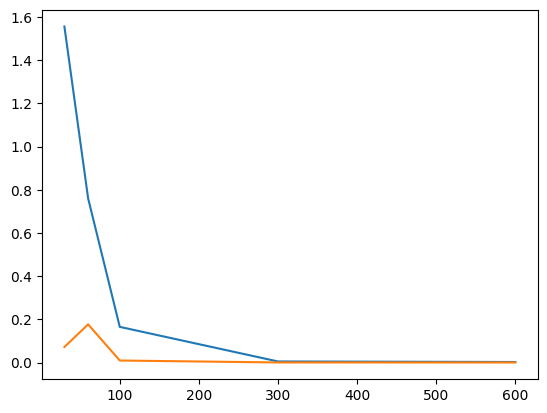
**RCC Model**

****

**Lowest Model**

****

**Mean Vs. Std**

****

F)

| **Model** | **Best M** | **Best G** | **D.O.F** | **No of Constraints** |
| --- | --- | --- | --- | --- |
| C | 3000 | 15 | 3000 | 4000 |
| D | 600 | 30 | 600 | 4000 |
| RCC D | 100 | 5 | 100 | 4000 |
| E | 600 | 15 | 600 | 4000 |
| RCC E | 100 | 5 | 100 | 4000 |

G)

| **Model** | **M** | **G** | **Test RMSE** |
| --- | --- | --- | --- |
| C | 3000 | 15 | 1.9472956416321626e-07 |
| D | 600 | 30 | 0.0034267597574372728 |
| RCC D | 100 | 5 | 0.19637052589152862 |
| E | 600 | 15 | 0.002324607659927773 |
| RCC E | 100 | 5 | 0.15401787961836907 |

H)When comparing the results from (b)-(g), we may notice performance differences for different values of M (or K) and lambda during model selection. Higher values of M (or K) and lower values of lambda result in better validation set performance but may lead to overfitting on the training set. Higher M (or K) values and lower lambda values can result in more complex models with a higher capacity to fit the training data but may generalize poorly to new data.

Lower values of M (or K) and higher values of lambda, on the other hand, tend to result in simpler models with a lower capacity to fit the training data but may generalize better to new data. As a result, we must strike a balance between model complexity and generalization performance, which can be accomplished through model selection techniques such as cross-validation.

I)

