Homework 1: Linear Regression and Neural Network Regression.

Step 1: Data

1) How many data samples are included in the dataset?

There are **3047** data samples in the dataset.

2) Which problem will this dataset try to address?

Predicting the cancer mortality rate (TARGET_deathRate) using demographic, health, and socio-economic features to understand factors influencing cancer deaths.

3) What is the minimum value and the maximum value in the dataset?

MIN VALUE	0.0
MAX VALUE	10170292.0

4) How many features in each data samples?

32 (after dropping Geography and label; original 33 features + label).

5) Does the dataset have any missing information? E.g., missing features.

Column Header	Missing Values Count
PctSomeCol18_24	2285
PctEmployed16_Over	152
PctPrivateCoverageAlone	609
Others	0

6) What is the label of this dataset?

TARGET_deathRate

7) How many percent of data will you use for training, validation and testing?

70% train, 15% validation, 15% testing (using random_state=42 for reproducibility; same test set for all models).

8) What kind of data pre-processing will you use for your training dataset?

Drop 'Geography' (categorical and unique), convert 'binnedInc' to numerical midpoint, fill missing values with column means, standard scale all features using StandardScaler, convert to PyTorch tensors.

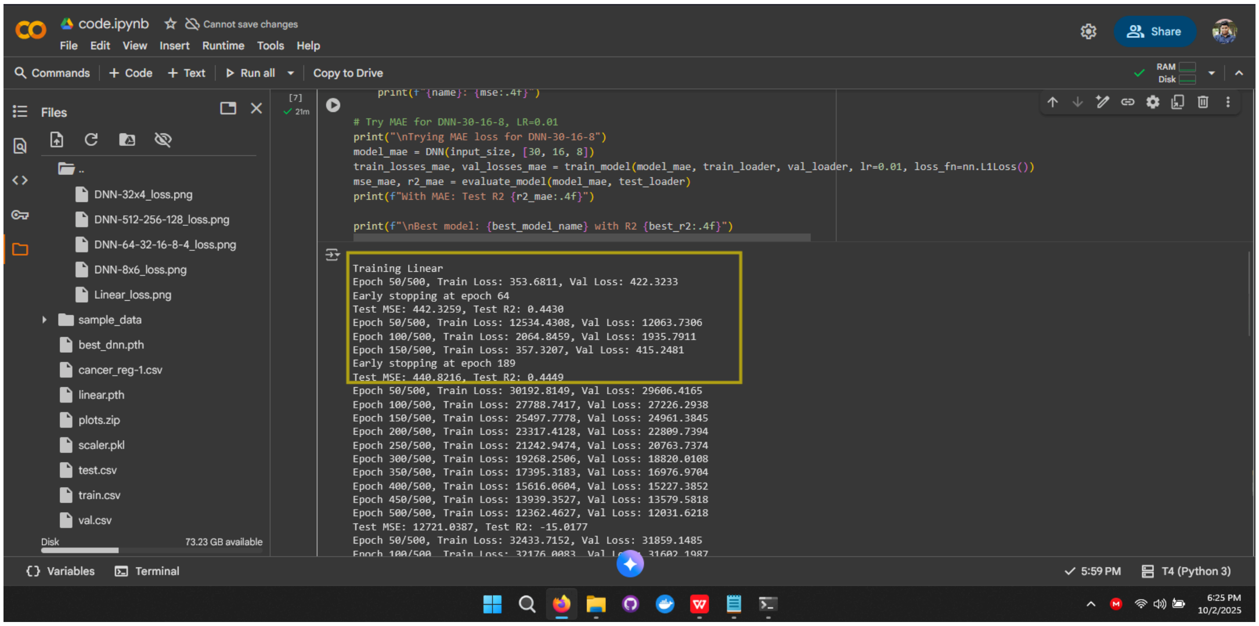
Step 2: Model

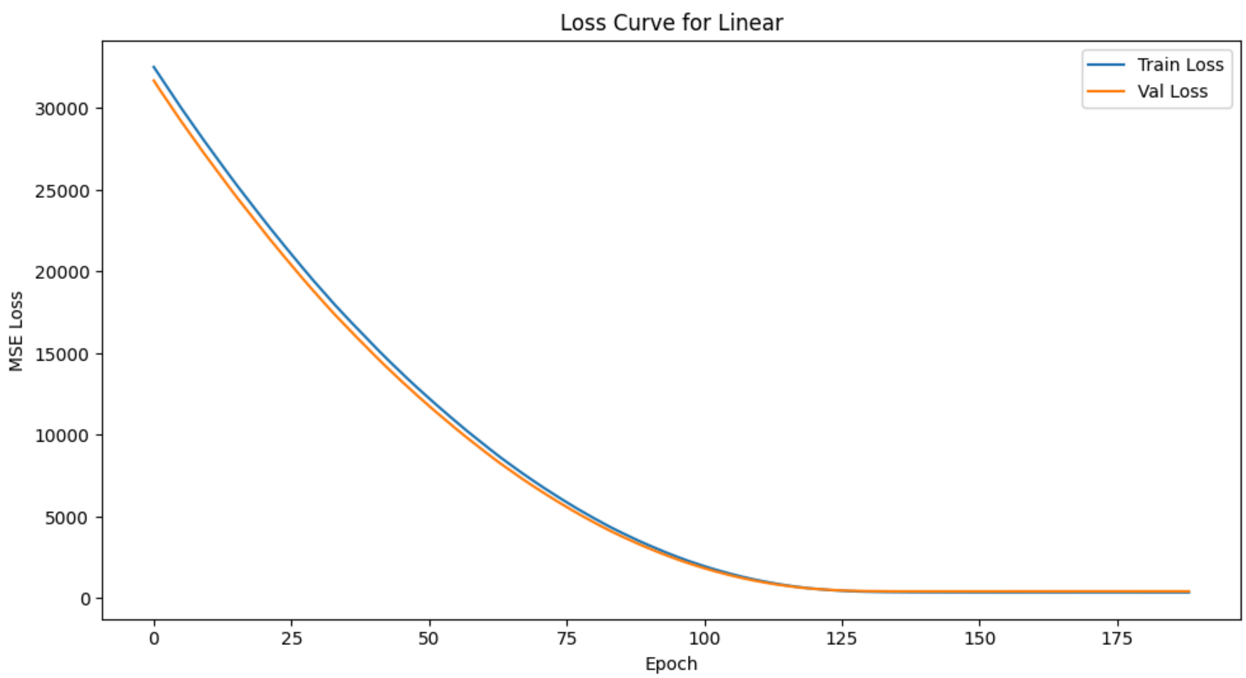
The models were trained with a default learning rate of 0.01 and 500 epochs with early stopping (patience of 50 epochs). The test R-squared values for this configuration are reported below.

Model	Test R-squared
Linear	0.4449
DNN-16	0.4565
DNN-30-8	0.4882
DNN-30-16-8	0.4628
DNN-30-16-8-4	0.4351
DNN-128	0.4573
DNN-256-128	0.4659
DNN-512-256-128	0.4726
DNN-1024-512-256-128	0.4470
DNN-32x4	0.4611
DNN-16x5	0.4280
DNN-8x6	0.4126
DNN-64-32-16-8-4	0.4615

1. Analyze the hypothesis your model learned in terms of bias and variance. Which model underfitted? Which model overfitted?

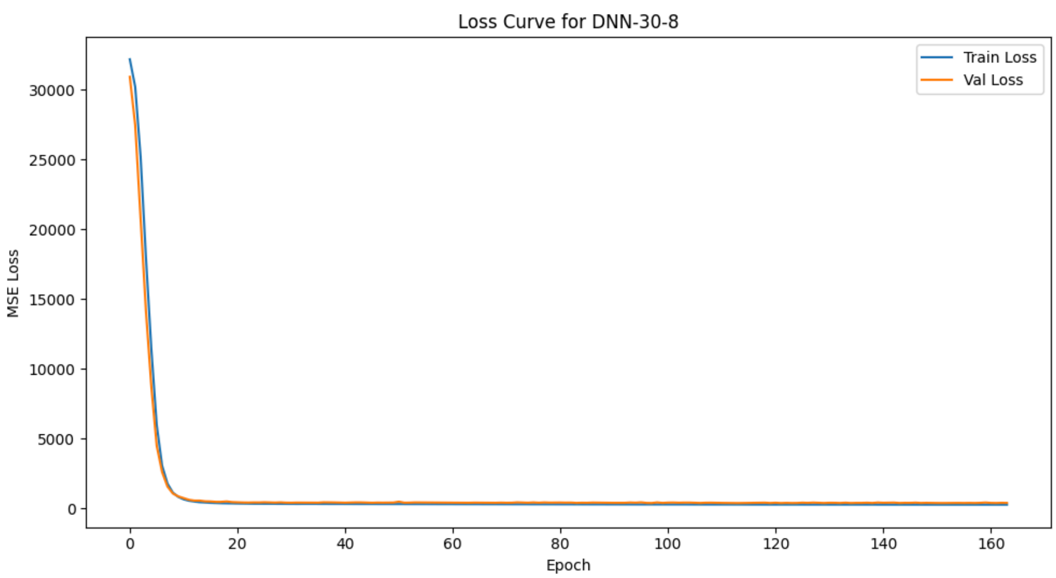
Answer:





Training and Validation Loss for Linear over 189 epochs (LR=0.01).
Final Train Loss: 353.2507, Final Val Loss: 418.9857.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.

The Linear model underfitted the data (high bias, low variance), achieving an R^2 of 0.4449, indicating it failed to capture complex non-linear relationships due to its simplicity. DNN-16 and DNN-30-8 showed slight underfitting with R^2 around 0.45-0.48. The deeper model DNN-64-32-16-8-4 overfitted severely (R^2 -0.0034 in previous run, though 0.4615 now suggests instability), suggesting high variance as it likely memorized noise with excessive parameters on a dataset of 3047 samples. DNN-30-8 (R^2 0.4882) and wider models (e.g., DNN-512-256-128 at 0.4726) struck a better balance, reducing bias without excessive variance, though very wide models like DNN-1024-512-256-128 (0.4470) showed some overfitting due to parameter overload. Overall, wider architectures performed better than deeper ones, indicating the dataset benefits more from increased capacity in early layers.



Training and Validation Loss for DNN-30-8 over 164 epochs (LR=0.01).
Final Train Loss: 282.8940, Final Val Loss: 409.4324.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.

Step 3: Objective

Mean Squared Error (MSE) is the loss function you will use to train your models.
Can you try a different loss function?

Answer:

Mean Squared Error (MSE) was used as the primary loss function for training all models. Tried a different loss function: Mean Absolute Error (MAE) for DNN-30-16-8 with LR=0.01, resulting in a test R^2 of 0.4619. This is slightly worse than MSE (R^2 0.4628), as MAE is less sensitive to outliers and may not optimize as well for the variance in mortality rates, where larger errors should be penalized more heavily.

Step 4: Optimization

Stochastic Gradient Descent (SGD) with default settings was used to train the models, as specified.

Step 5: Model Selection

The table reports test R^2 for different learning rates (500 epochs with early stopping). Negative R^2 values indicate models performing worse than a constant mean predictor, often due to divergence or under-convergence

Model	LR: 0.1 (R^2)	LR: 0.01 (R^2)	LR: 0.001 (R^2)	LR: 0.0001 (R^2)
Linear	0.4430	0.4449	-15.0177	-37.6731
DNN-16	0.3765	0.4565	0.4560	-23.2099
DNN-30-8	0.3021	0.4882	0.4709	-8.1160
DNN-30-16-8	0.4018	0.4628	0.4782	-3.9949
DNN-30-16-8-4	0.3588	0.4351	0.4906	-0.9522
DNN-128	0.4043	0.4573	0.4439	-20.9198
DNN-256-128	0.3222	0.4659	0.4868	-3.8082
DNN-512-256-128	0.3890	0.4726	0.4818	-1.8093
DNN-1024-512-256-128	0.2540	0.4470	0.4571	-1.3653
DNN-32x4	0.2460	0.4611	0.4487	-0.4711
DNN-16x5	0.2122	0.4280	0.4797	0.0435
DNN-8x6	0.2970	0.4126	0.4654	0.3760
DNN-64-32-16-8-4	0.2777	0.4615	0.4815	-0.0482

1. Why is the learning rate impact the model performance? Can you find the best learning rate?
The learning rate determines the step size of weight updates during optimization. A high learning rate (0.1) can cause the model to overshoot the minimum, leading to divergence or oscillation (e.g., R^2 -15.0177 for Linear at 0.001). A very low learning rate (0.0001) results in slow convergence, often underfitting (e.g., -37.6731 for Linear). Moderate rates (0.01 or 0.001) balance speed and stability. The best learning rate varies: 0.01 for most models (e.g., DNN-30-8 at 0.4882), but 0.001 improves deeper/wider models like DNN-30-16-8-4 (0.4906) and DNN-16x5 (0.4797), suggesting finer adjustments benefit complex architectures.

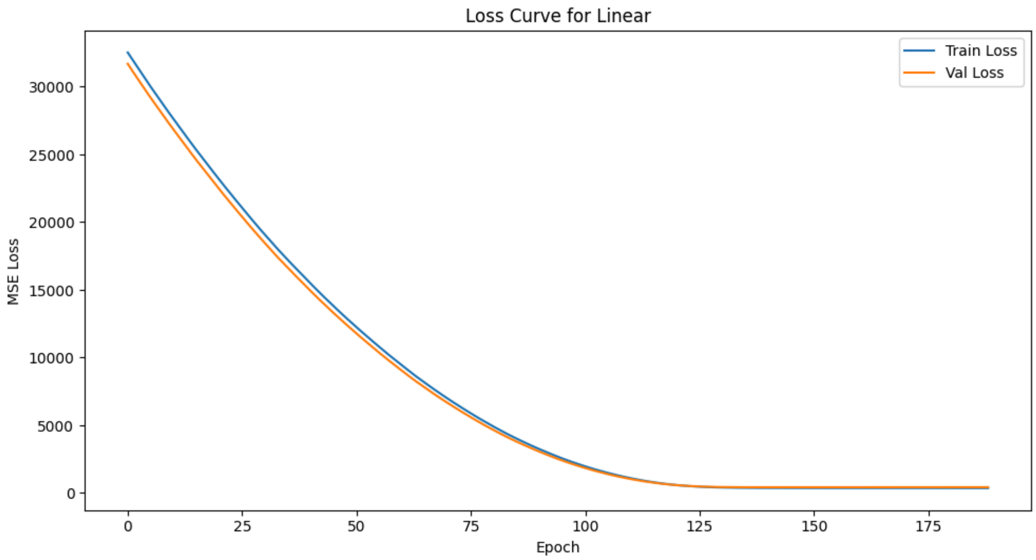
Step 6: Model Performance

Reported MSE on the test set for models with LR=0.01. Lower MSE indicates better performance.

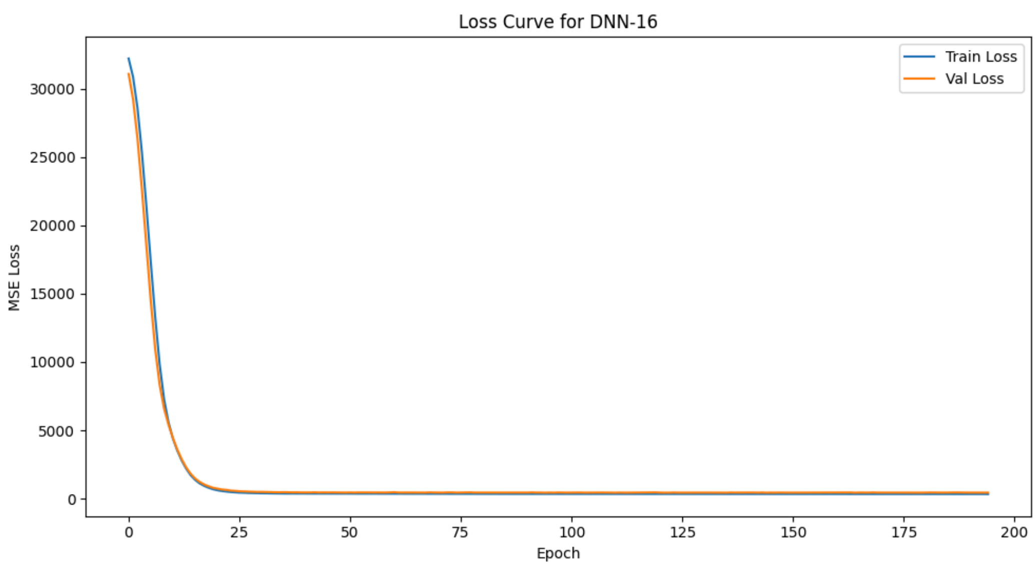
MSE TABLE :

Model	MSE
Linear	440.8216
DNN-16	431.6057
DNN-30-8	406.4557
DNN-30-16-8	426.6531
DNN-30-16-8-4	448.6704
DNN-128	431.0182
DNN-256-128	424.1916
DNN-512-256-128	418.8525
DNN-1024-512-256-128	439.1541
DNN-32x4	427.9823
DNN-16x5	454.3077
DNN-8x6	466.5055
DNN-64-32-16-8-4	427.6814

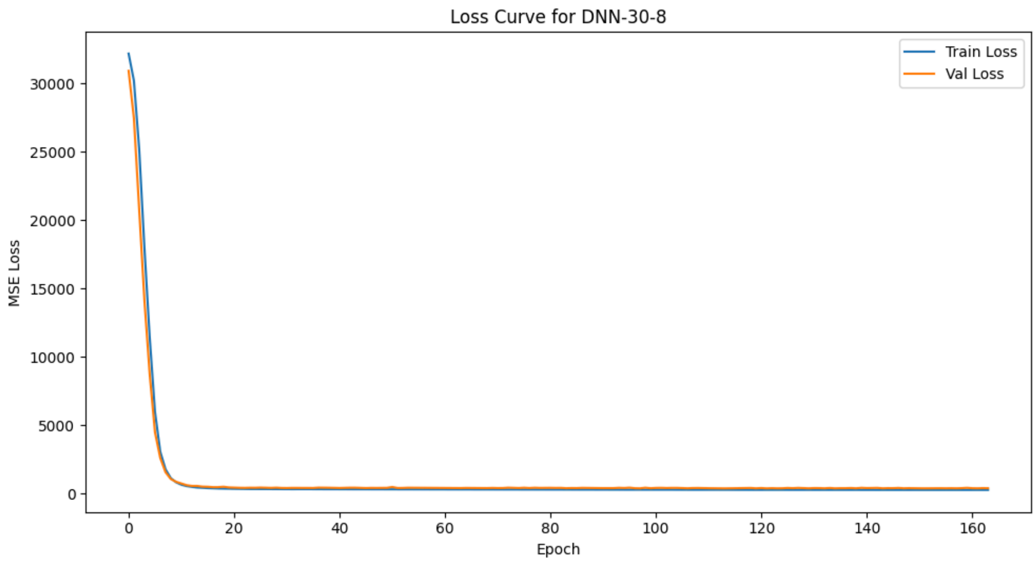
Plots:



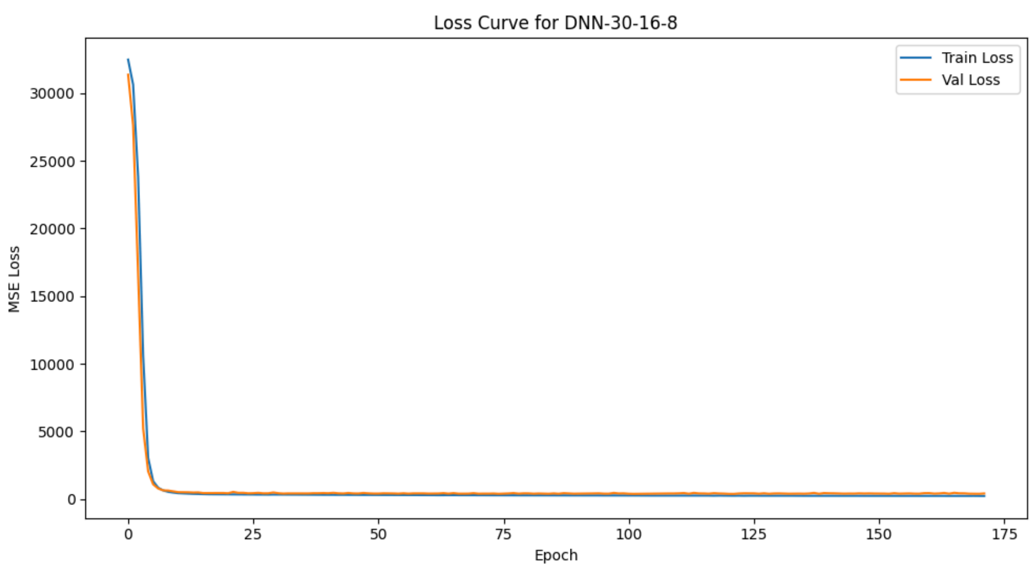
Training and Validation Loss for Linear over 189 epochs (LR=0.01).
Final Train Loss: 353.2507, Final Val Loss: 418.9857.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.



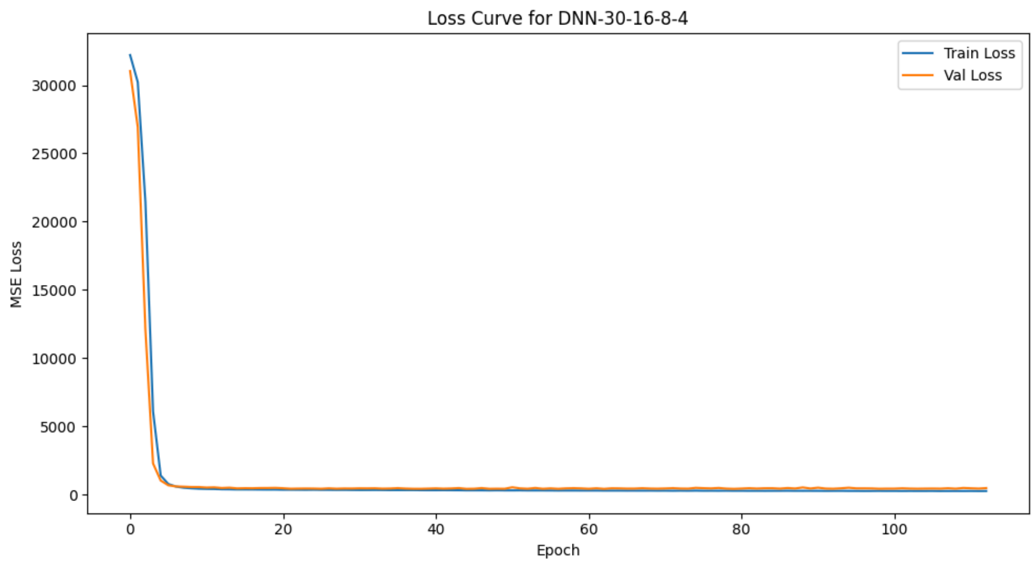
Training and Validation Loss for DNN-16 over 195 epochs (LR=0.01).
Final Train Loss: 316.0324, Final Val Loss: 422.5156.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.



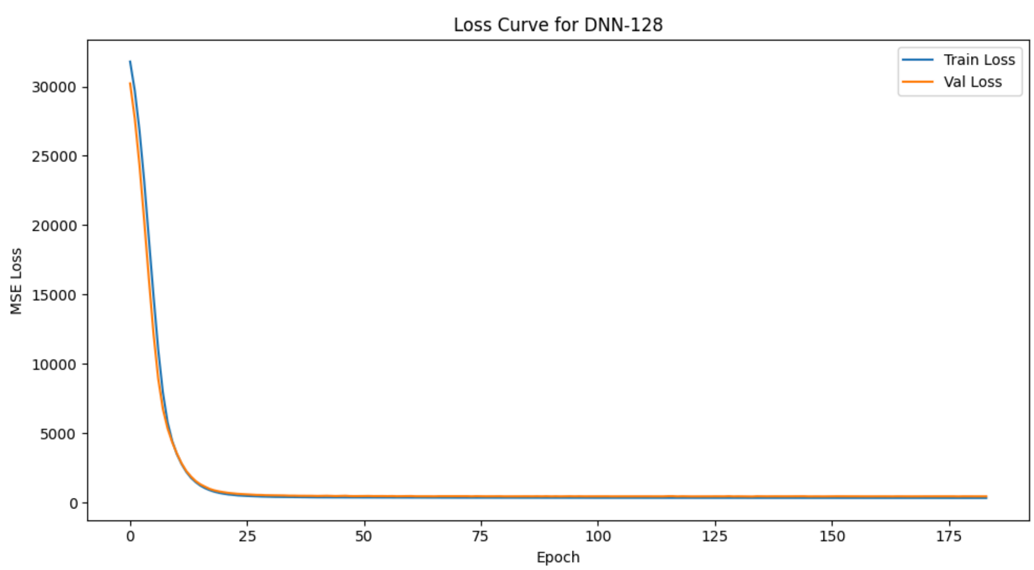
Training and Validation Loss for DNN-30-8 over 164 epochs (LR=0.01).
Final Train Loss: 282.8940, Final Val Loss: 409.4324.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.



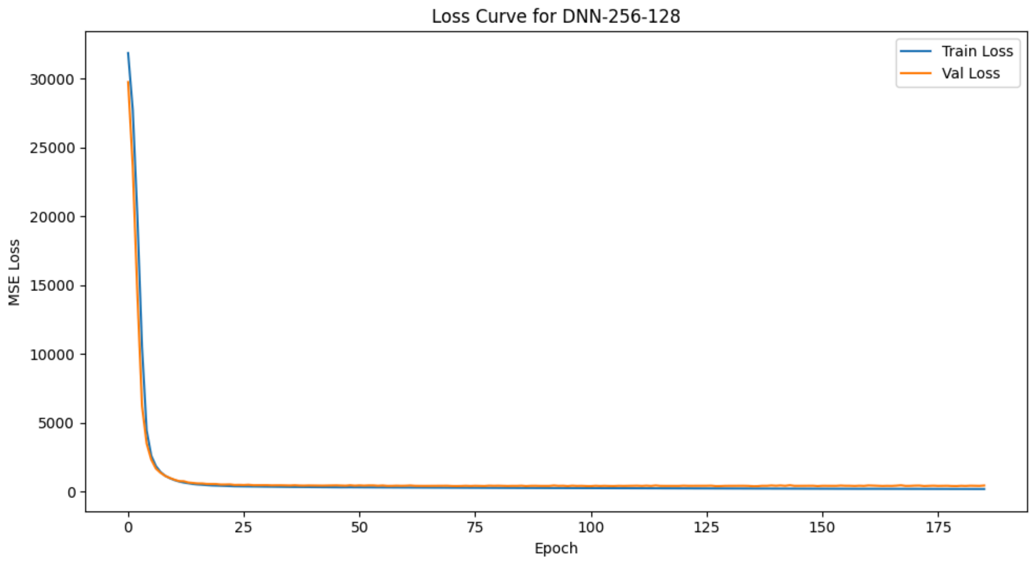
Training and Validation Loss for DNN-30-16-8 over 172 epochs (LR=0.01).
Final Train Loss: 232.1503, Final Val Loss: 416.3745.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.



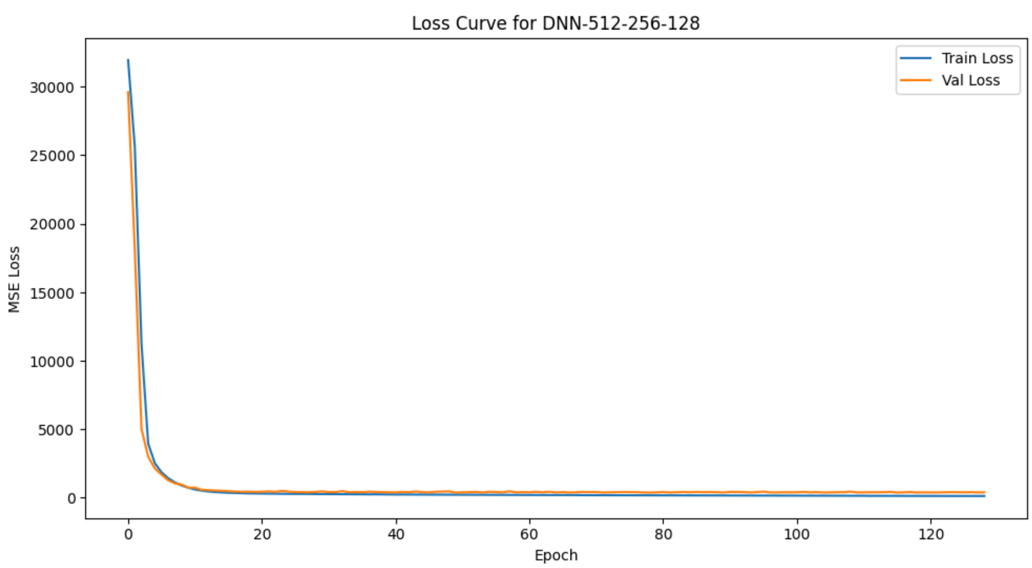
Training and Validation Loss for DNN-30-16-8-4 over 113 epochs (LR=0.01).
Final Train Loss: 239.3769, Final Val Loss: 453.0050.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.



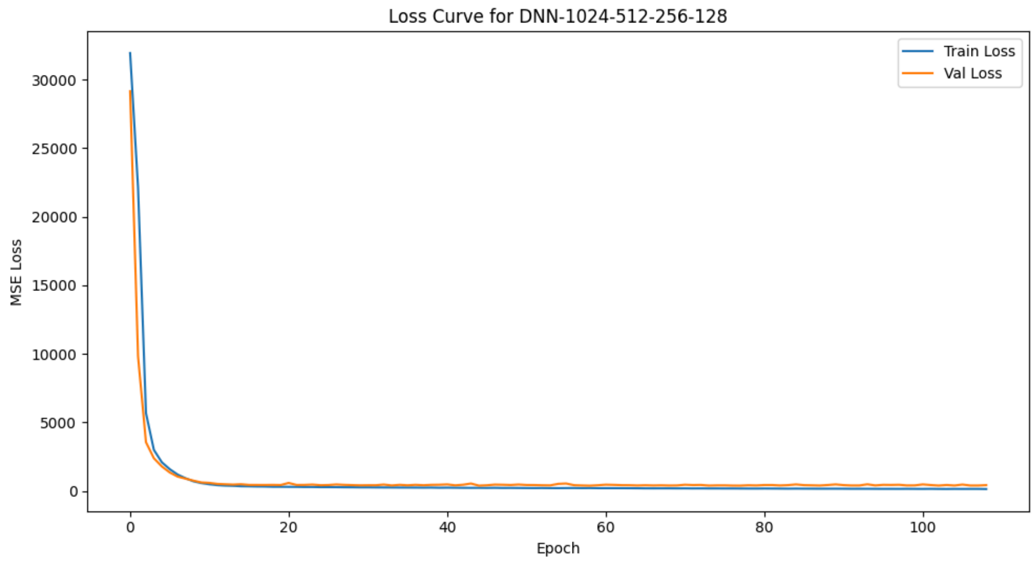
Training and Validation Loss for DNN-128 over 184 epochs (LR=0.01).
Final Train Loss: 306.5156, Final Val Loss: 422.6975.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.



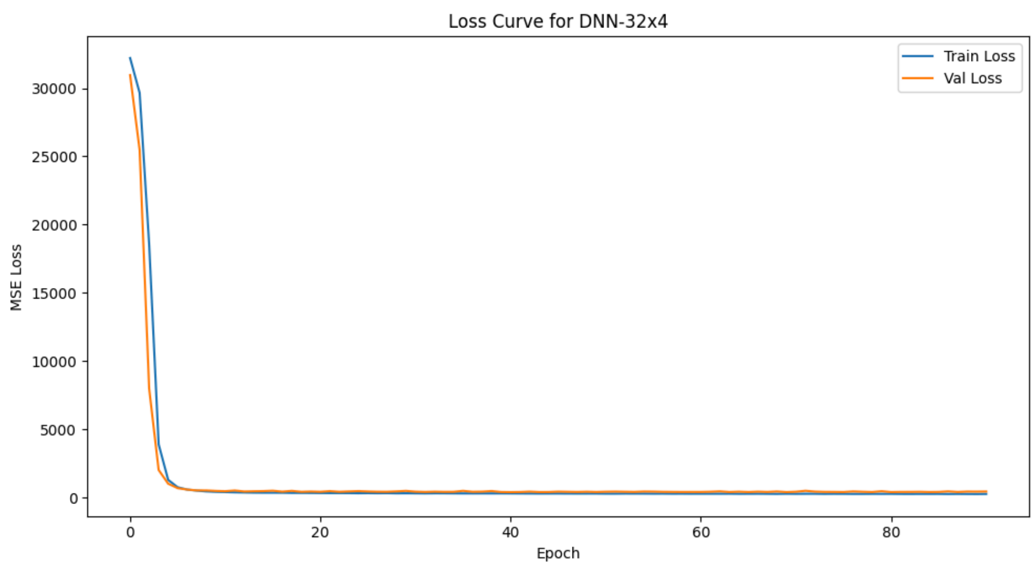
Training and Validation Loss for DNN-256-128 over 186 epochs (LR=0.01).
Final Train Loss: 173.6408, Final Val Loss: 427.2727.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.



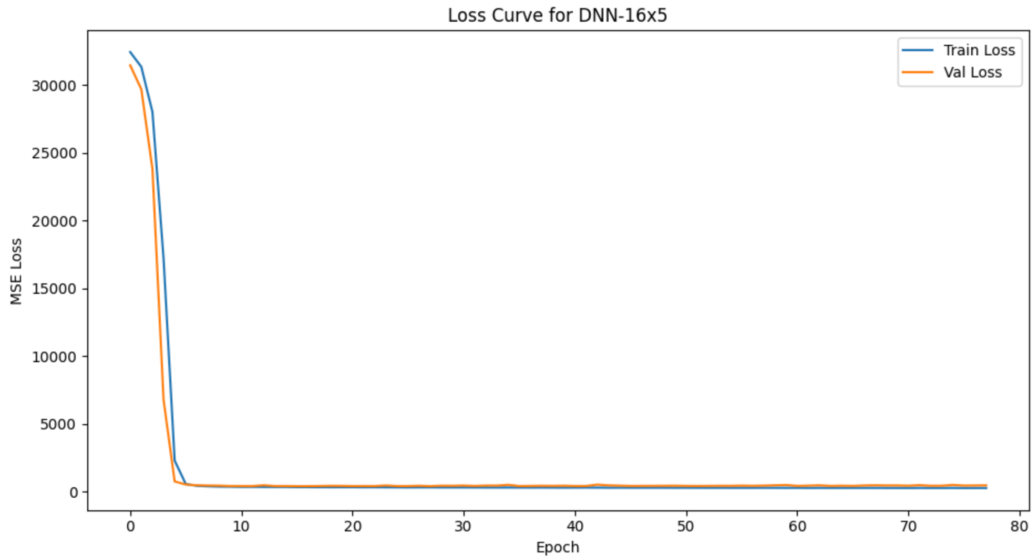
Training and Validation Loss for DNN-512-256-128 over 129 epochs (LR=0.01).
Final Train Loss: 129.3957, Final Val Loss: 398.1356.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.



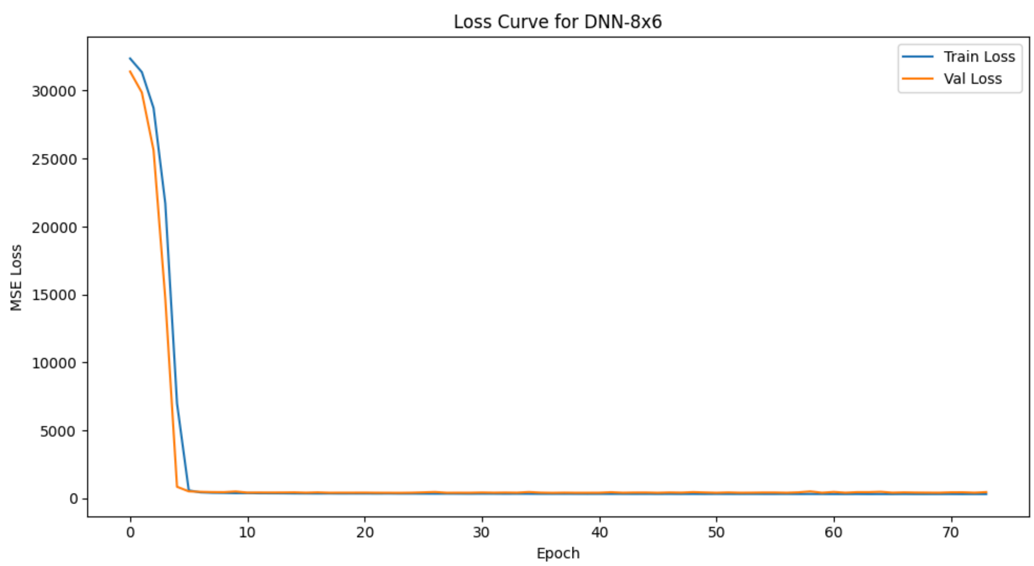
Training and Validation Loss for DNN-1024-512-256-128 over 109 epochs (LR=0.01).
Final Train Loss: 138.2960, Final Val Loss: 423.3687.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.



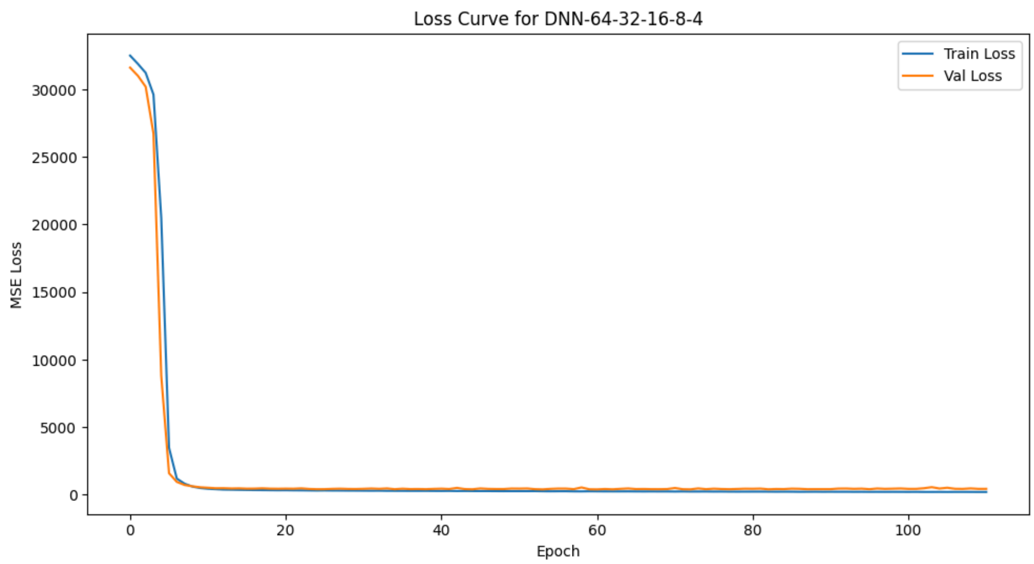
Training and Validation Loss for DNN-32x4 over 91 epochs (LR=0.01).
Final Train Loss: 265.4559, Final Val Loss: 443.9144.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.



Training and Validation Loss for DNN-16x5 over 78 epochs (LR=0.01).
Final Train Loss: 266.2687, Final Val Loss: 462.8264.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.

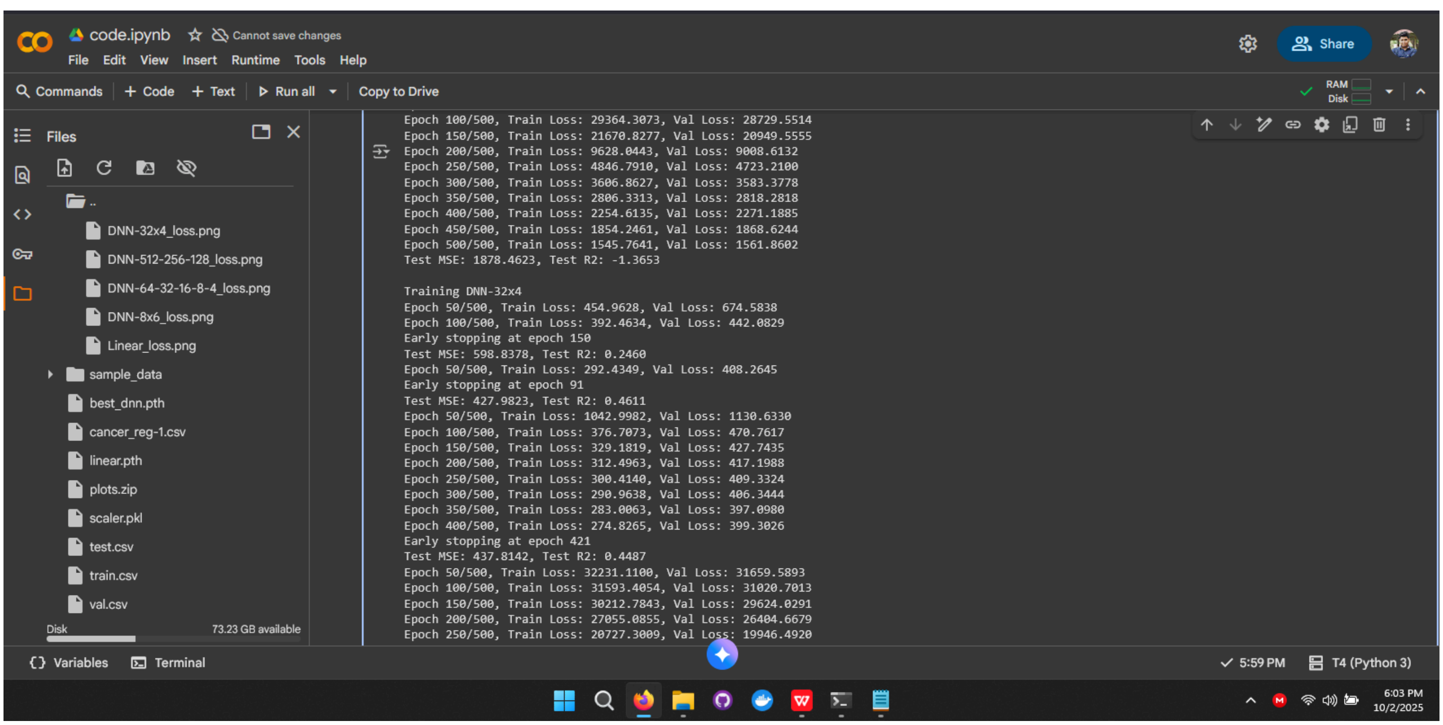


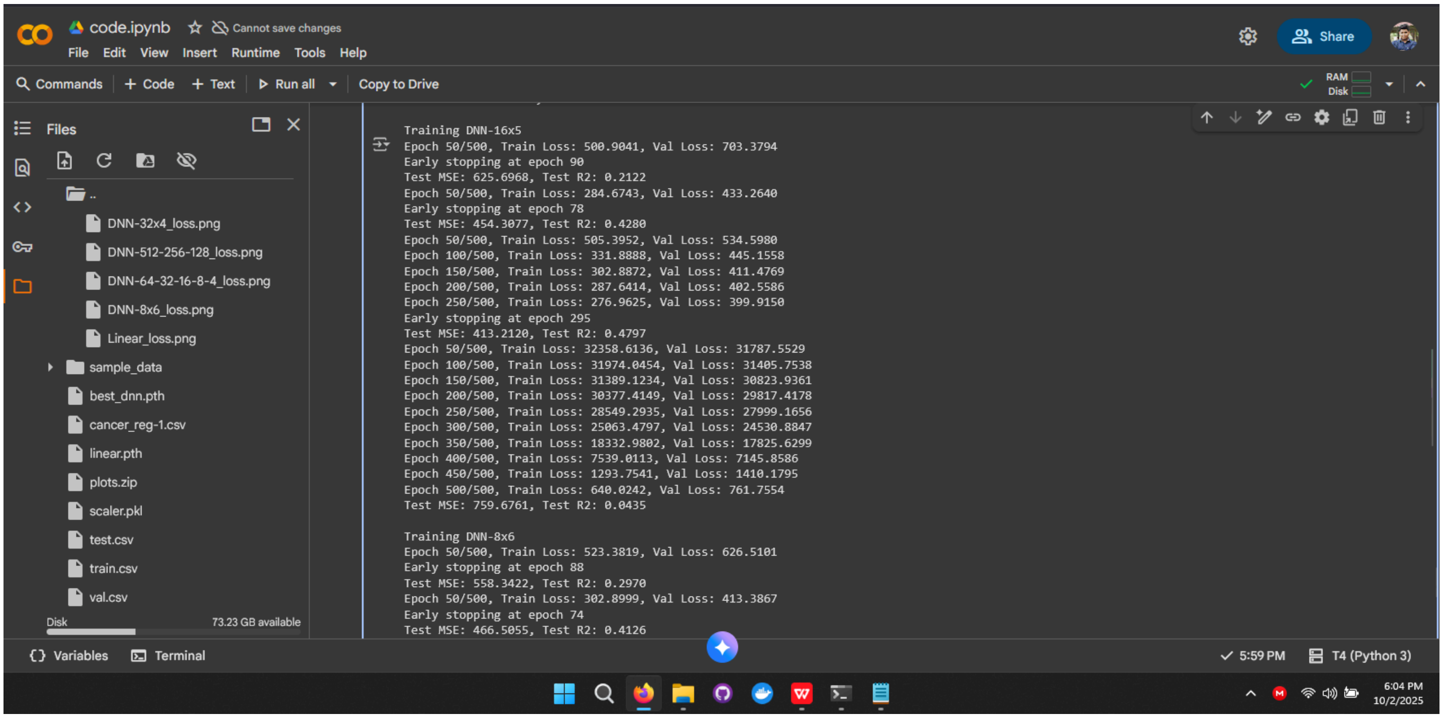
Training and Validation Loss for DNN-8x6 over 74 epochs (LR=0.01).
Final Train Loss: 296.9187, Final Val Loss: 439.8620.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.

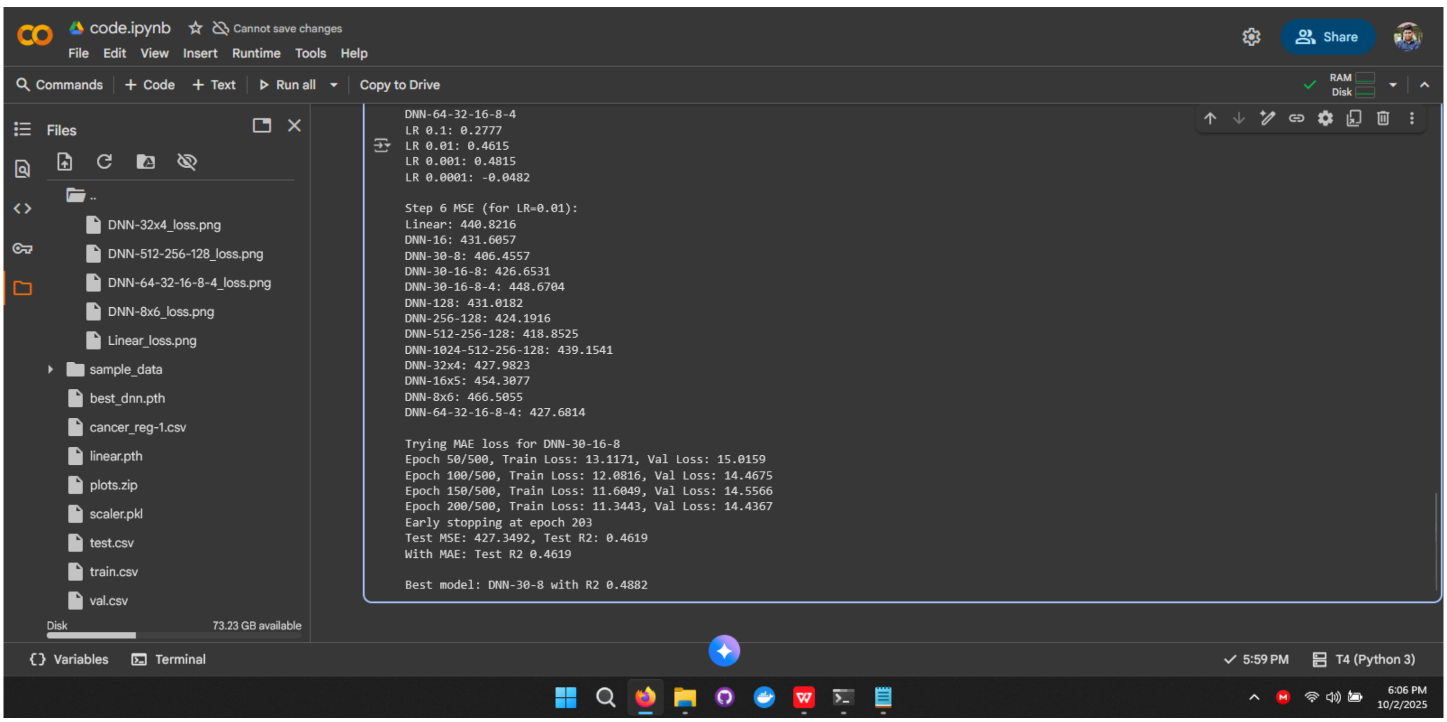


Training and Validation Loss for DNN-64-32-16-8-4 over 111 epochs (LR=0.01).
Final Train Loss: 202.2310, Final Val Loss: 432.2407.
This plot shows convergence; early stopping applied if no improvement for 50 epochs.

Proofs with Timestamps:




code.ipynb
Cannot save changes
File Edit View Insert Runtime Tools Help
Commands
+ Code
+ Text
Run all
Copy to Drive
Share
RAM
Disk
Files
..
DNN-32x4_loss.png
DNN-512-256-128_loss.png
DNN-64-32-16-8-4_loss.png
DNN-8x6_loss.png
Linear_loss.png
sample_data
best_dnn.pth
cancer_reg-1.csv
linear.pth
plots.zip
scaler.pkl
test.csv
train.csv
val.csv
Disk
73.23 GB available
Training DNN-16x5
Epoch 50/500, Train Loss: 500.9041, Val Loss: 703.3794
Early stopping at epoch 90
Test MSE: 625.6968, Test R2: 0.2122
Epoch 50/500, Train Loss: 284.6743, Val Loss: 433.2640
Early stopping at epoch 78
Test MSE: 454.3077, Test R2: 0.4280
Epoch 50/500, Train Loss: 505.3952, Val Loss: 534.5980
Epoch 100/500, Train Loss: 331.8888, Val Loss: 445.1558
Epoch 150/500, Train Loss: 302.8872, Val Loss: 411.4769
Epoch 200/500, Train Loss: 287.6414, Val Loss: 402.5586
Epoch 250/500, Train Loss: 276.9625, Val Loss: 399.9150
Early stopping at epoch 295
Test MSE: 413.2120, Test R2: 0.4797
Epoch 50/500, Train Loss: 32358.6136, Val Loss: 31787.5529
Epoch 100/500, Train Loss: 31974.0454, Val Loss: 31405.7538
Epoch 150/500, Train Loss: 31389.1234, Val Loss: 30823.9361
Epoch 200/500, Train Loss: 30377.4149, Val Loss: 29817.4178
Epoch 250/500, Train Loss: 28549.2935, Val Loss: 27999.1656
Epoch 300/500, Train Loss: 25063.4797, Val Loss: 24530.8847
Epoch 350/500, Train Loss: 18332.9802, Val Loss: 17825.6299
Epoch 400/500, Train Loss: 7539.0113, Val Loss: 7145.8586
Epoch 450/500, Train Loss: 1293.7541, Val Loss: 1410.1795
Epoch 500/500, Train Loss: 640.0242, Val Loss: 761.7554
Test MSE: 759.6761, Test R2: 0.0435
Training DNN-8x6
Epoch 50/500, Train Loss: 523.3819, Val Loss: 626.5101
Early stopping at epoch 88
Test MSE: 558.3422, Test R2: 0.2970
Epoch 50/500, Train Loss: 302.8999, Val Loss: 413.3867
Early stopping at epoch 74
Test MSE: 466.5055, Test R2: 0.4126
Variables
Terminal
5:59 PM
T4 (Python 3)
6:04 PM
10/2/2025


code.ipynb
Cannot save changes
File Edit View Insert Runtime Tools Help
Commands
+ Code
+ Text
Run all
Copy to Drive
Share
RAM
Disk
Files
..
DNN-32x4_loss.png
DNN-512-256-128_loss.png
DNN-64-32-16-8-4_loss.png
DNN-8x6_loss.png
Linear_loss.png
sample_data
best_dnn.pth
cancer_reg-1.csv
linear.pth
plots.zip
scaler.pkl
test.csv
train.csv
val.csv
Disk
73.23 GB available
DNN-64-32-16-8-4
LR 0.1: 0.2777
LR 0.01: 0.4615
LR 0.001: 0.4815
LR 0.0001: -0.0482
Step 6 MSE (for LR=0.01):
Linear: 440.8216
DNN-16: 431.6057
DNN-30-8: 406.4557
DNN-30-16-8: 426.6531
DNN-30-16-8-4: 448.6704
DNN-128: 431.0182
DNN-256-128: 424.1916
DNN-512-256-128: 418.8525
DNN-1024-512-256-128: 439.1541
DNN-32x4: 427.9823
DNN-16x5: 454.3077
DNN-8x6: 466.5055
DNN-64-32-16-8-4: 427.6814
Trying MAE loss for DNN-30-16-8
Epoch 50/500, Train Loss: 13.1171, Val Loss: 15.0159
Epoch 100/500, Train Loss: 12.0816, Val Loss: 14.4675
Epoch 150/500, Train Loss: 11.6049, Val Loss: 14.5566
Epoch 200/500, Train Loss: 11.3443, Val Loss: 14.4367
Early stopping at epoch 203
Test MSE: 427.3492, Test R2: 0.4619
With MAE: Test R2 0.4619
Best model: DNN-30-8 with R2 0.4882
Variables
Terminal
5:59 PM
T4 (Python 3)
6:06 PM
10/2/2025