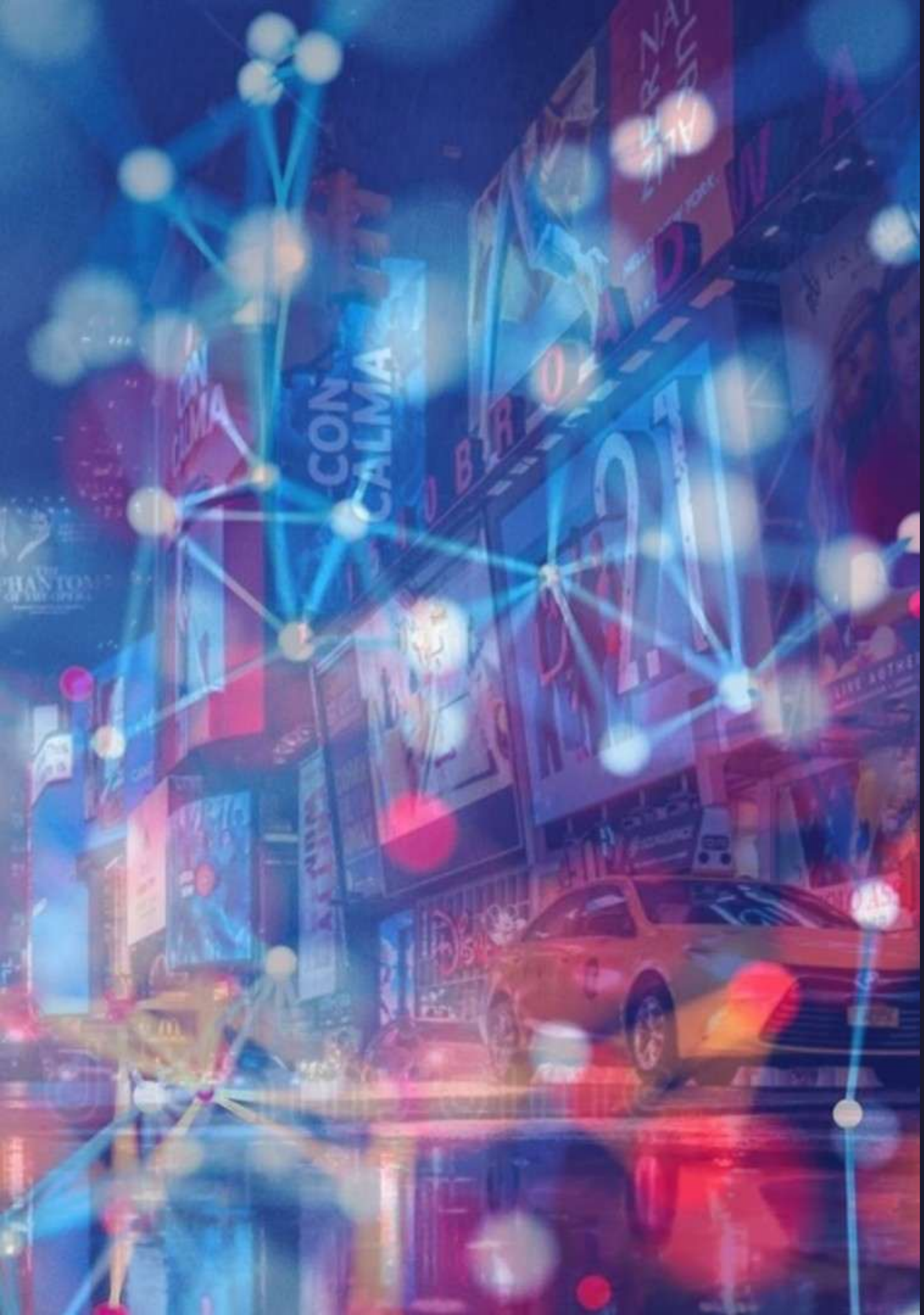


Course: CS672 – Introduction to Deep Learning

Project: #2

Deep Learning for NYC Taxi Trip Duration Prediction

- Team: Krishna Pothana & Xianrong Liang
- Professor: Tassos Sarbanes
- Date: Nov 7, 2025

Agenda

- Introduction (Problem & Goal)
- Feature Engineering (Data Processing)
- Model Design(TensorFlow)
- PyTorch Version
- Results Comparison
- Conclusion
- Code Showcase

Introduction (Problem & Goal)

Background

- Predict the trip duration of NYC yellow cabs
- Data sources:
  - NYC Taxi Dataset (01/01/2020~01/31/2020)
  - Meteostat Weather Data (Wall Street, NYC)

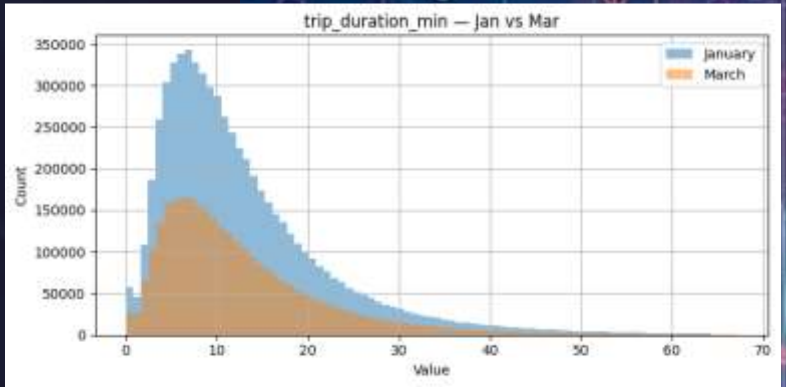
Goal

- Build a regression model based on neural networks
- Improve prediction accuracy by feature integration (in combination with meteorological characteristics)

Feature Engineering (Data Processing)

Processing Steps

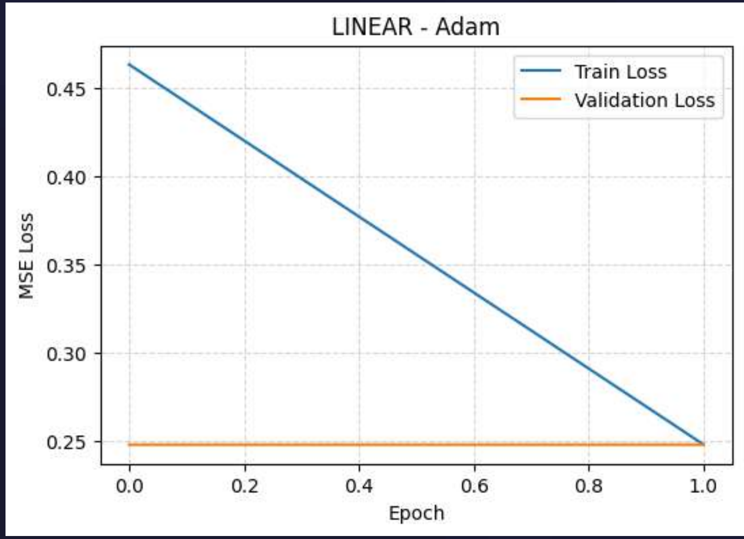
1. Remove outliers (duration ≤ 0 or > 180 minutes)
2. Extract time features: `hour`, `day_of_week`
3. Integrated weather features: `tavg`, `prcp`, `wspd`
4. Normalization: `log1p` (trip_duration)
5. Standardization (StandardScaler)

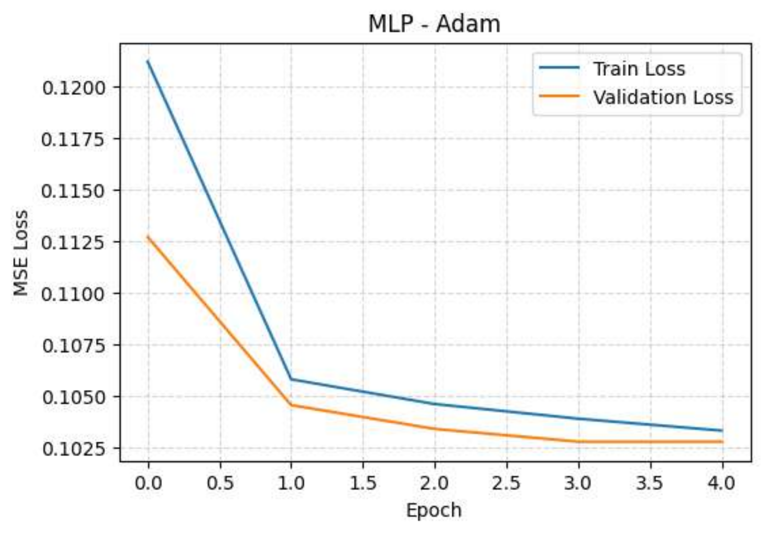


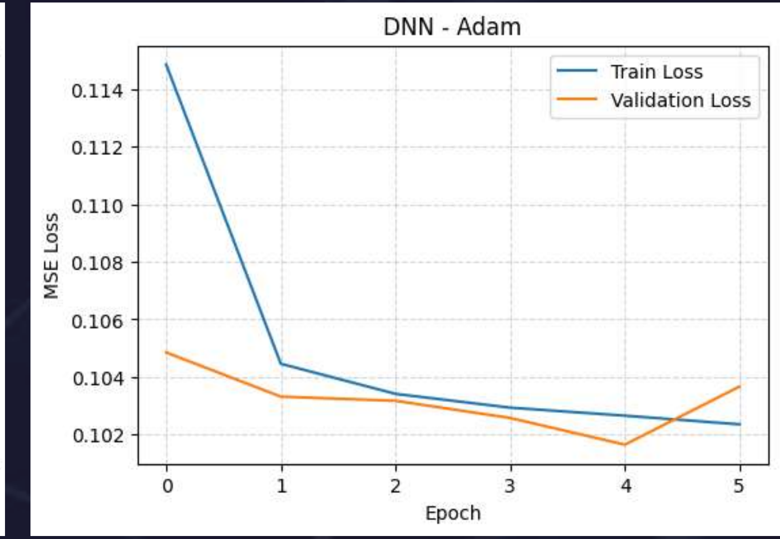
Model Design (TensorFlow / Keras)

Types of Model	Hidden layer structure	Optimizer	Feature
Linear Regression	No hidden layer	Adam	Baseline Model
MLP (Multi-Layer Perceptron)	64–32	Adam / RMSProp	Simple nonlinearity
DNN(Deep Neural Network)	128–64–32	Adam	Optimal performance

Model Design (+Adam)

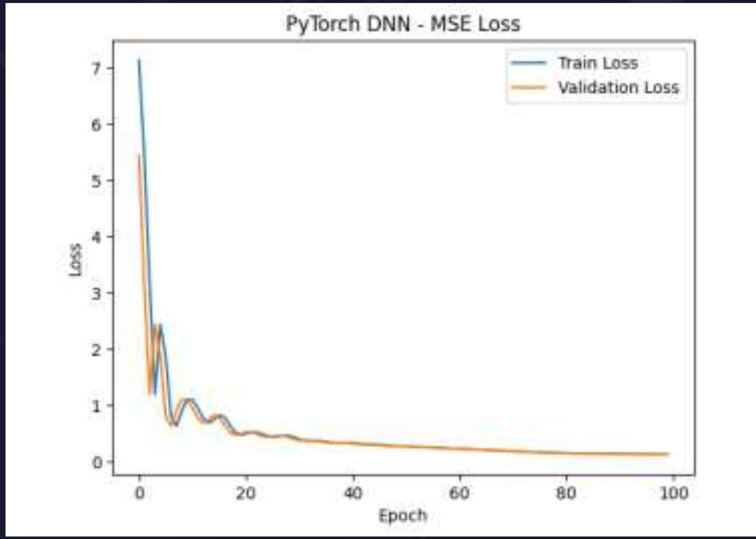




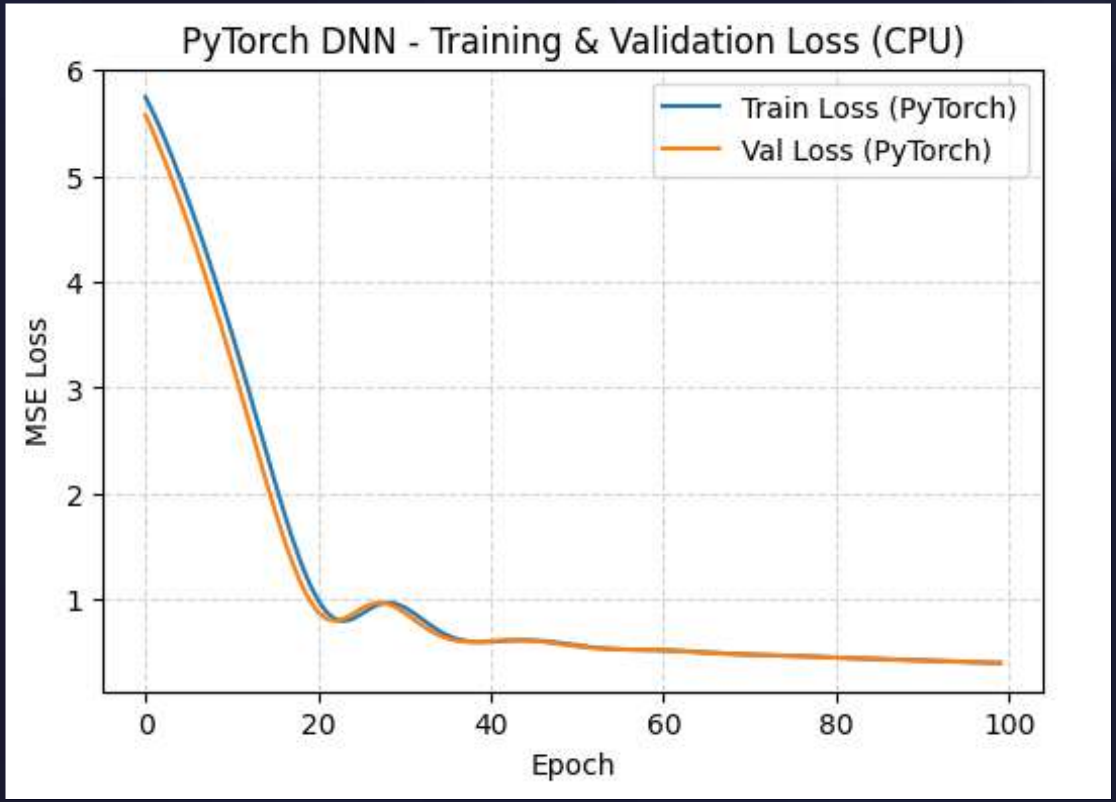


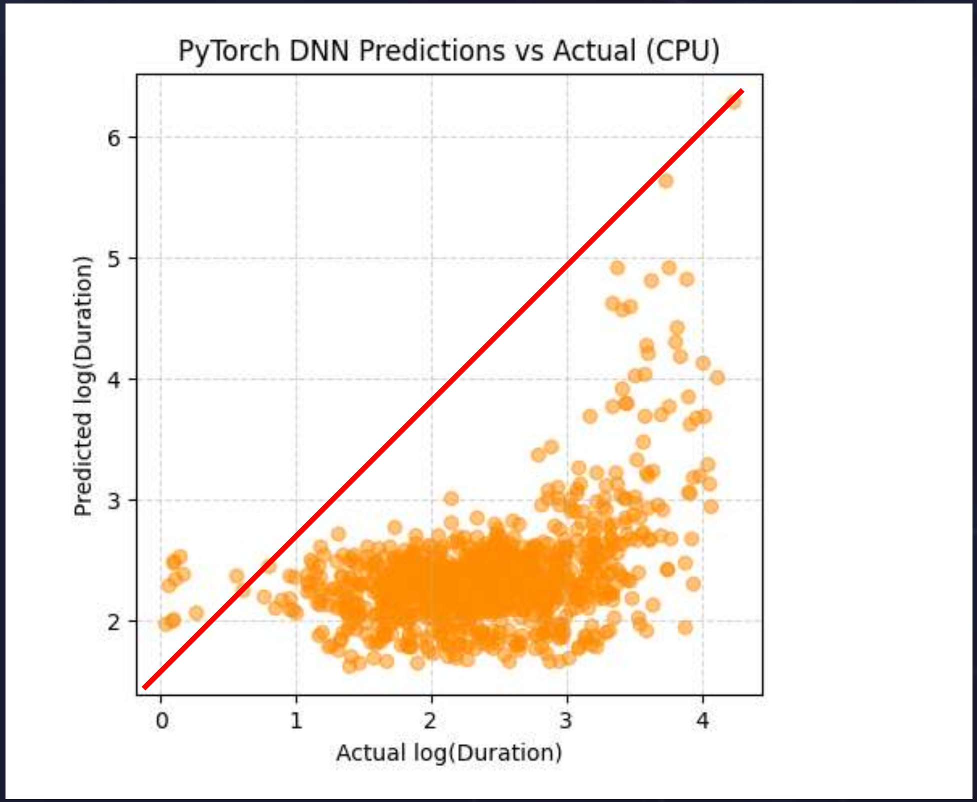
Linear Regression vs. MLP vs. DNN vs. PyTorch DNN

Train Loss Curve vs. Validate the Loss curve



PyTorch Version DNN





Reproduce the DNN model in PyTorch

Use GPU acceleration for training

The verification result is close to TensorFlow

Prediction vs. Actual scatter plot

Findings and analysis

Model	Val MSE	Val MAE	Framework
Linear	0.248185	0.381835	TF
MLP	0.102788	0.226832	TF
DNN	0.101652	0.224575	TF
DNN (PyTorch-CPU)	0.10296	0.225242	PT

Best Model: DNN | Optimizer: Adam | MAE: 0.2246

Key points of the conclusion

- Improve the accuracy of predictions
- DNN has the best generalization and avoids overfitting
- Comparable to that of TF, which verifies the stability of PyTorch

Conclusion & Future Work

- Deep learning can effectively model the nonlinear relationship of travel time
- The integration of weather data significantly enhances the prediction effect
- Future Expectations：
  - Time series modeling is conducted using Transformer TensorFlow
  - Add real-time traffic data
  - Predict multi-day trends

Course: CS672 – Introduction to Deep Learning

Code Showcase

Course: CS672 – Introduction to Deep Learning

Q&A / Thank you

Project #2: Deep Learning for NYC Taxi Trip Duration Prediction

Team #4: Krishna Pothana & Xianrong Liang