

Statistics and Data Analysis

Unit 04 – Lecture 09: Case Study: End-to-End Regression Workflow

Tofik Ali

School of Computer Science, UPES Dehradun

February 14, 2026

<https://github.com/tali7c/Statistics-and-Data-Analysis>

Overview
○

Workflow
○

Evaluation
○

Exercises
○○○○○

Demo
○○

Summary
○○

Quick Links

Overview

Workflow

Evaluation

Exercises

Demo

Summary

Agenda

1 Overview

2 Workflow

3 Evaluation

4 Exercises

5 Demo

6 Summary

Learning Outcomes

- Describe an end-to-end regression workflow

Learning Outcomes

- Describe an end-to-end regression workflow
- Choose appropriate regression metrics (RMSE and R^2)

Learning Outcomes

- Describe an end-to-end regression workflow
- Choose appropriate regression metrics (RMSE and R^2)
- Check overfitting (train vs test gap)

Learning Outcomes

- Describe an end-to-end regression workflow
- Choose appropriate regression metrics (RMSE and R^2)
- Check overfitting (train vs test gap)
- Communicate results with plots (predicted vs actual, residuals)

Workflow: Key Points

- Define target and inputs

Workflow: Key Points

- Define target and inputs
- Prepare data and split chronologically if needed

Workflow: Key Points

- Define target and inputs
- Prepare data and split chronologically if needed
- Fit baseline then iterate

Evaluation: Key Points

- Use RMSE/MSE/MAE and R^2

Evaluation: Key Points

- Use RMSE/MSE/MAE and R^2
- Use plots: predicted vs actual, residuals

Evaluation: Key Points

- Use RMSE/MSE/MAE and R^2
- Use plots: predicted vs actual, residuals
- Document limitations

Exercise 1: Metric choice

Target is continuous (price). Should you use accuracy?

Solution 1

- No; accuracy is for classification.

Exercise 2: Overfitting sign

Train RMSE=5, test RMSE=20. What does it suggest?

Solution 2

- Overfitting; try simpler model or regularization.

Exercise 3: Communication

Name one plot to communicate regression quality.

Solution 3

- Predicted vs actual scatter; residual plot.

Mini Demo (Python)

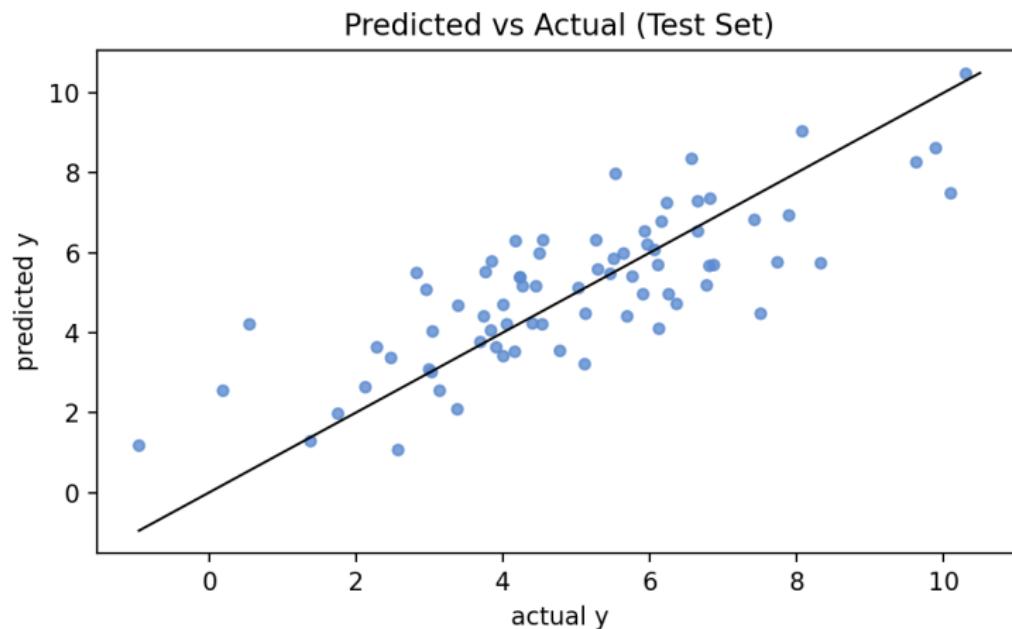
Run from the lecture folder:

```
python demo/demo.py
```

Outputs:

- images/demo.png
- data/results.txt

Demo Output (Example)



Summary

- Key definitions and the main formula.

Summary

- Key definitions and the main formula.
- How to interpret results in context.

Summary

- Key definitions and the main formula.
- How to interpret results in context.
- How the demo connects to the theory.

Exit Question

What would you do first if the case study model performs poorly on the test set?