

Statistics and Data Analysis

Unit 05 – Lecture 06: Advanced Feature Engineering for Multivariate Data

Tofik Ali

School of Computer Science, UPES Dehradun

February 17, 2026

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

Quick Links

Overview

Interactions

Aggregations

Exercises

Demo

Summary

Agenda

- 1 Overview
- 2 Interactions
- 3 Aggregations
- 4 Exercises
- 5 Demo
- 6 Summary

Learning Outcomes

- Create interaction features when meaningful

Learning Outcomes

- Create interaction features when meaningful
- Create aggregation features from transactional data

Learning Outcomes

- Create interaction features when meaningful
- Create aggregation features from transactional data
- Engineer time-based features (lags/rolling)

Learning Outcomes

- Create interaction features when meaningful
- Create aggregation features from transactional data
- Engineer time-based features (lags/rolling)
- Avoid leakage and look-ahead bias

Interactions: Key Points

- Products and ratios capture combined effects

Interactions: Key Points

- Products and ratios capture combined effects
- Use domain knowledge

Interactions: Key Points

- Products and ratios capture combined effects
- Use domain knowledge
- Validate with CV

Aggregations: Key Points

- Per-user totals/means/counts

Aggregations: Key Points

- Per-user totals/means/counts
- Rolling windows (last 7/30 days)

Aggregations: Key Points

- Per-user totals/means/counts
- Rolling windows (last 7/30 days)
- Avoid using future data

Exercise 1: Interaction

Give one interaction feature for house price.

Solution 1

■ `size_m2 * location_score (example).`

Exercise 2: Aggregation

Name one per-user aggregation for churn prediction.

Solution 2

- `days_since_last_purchase` (example).

Exercise 3: Leakage

Is using next-30-days spend to predict churn today leakage?

Solution 3

- Yes; it uses future info.

Mini Demo (Python)

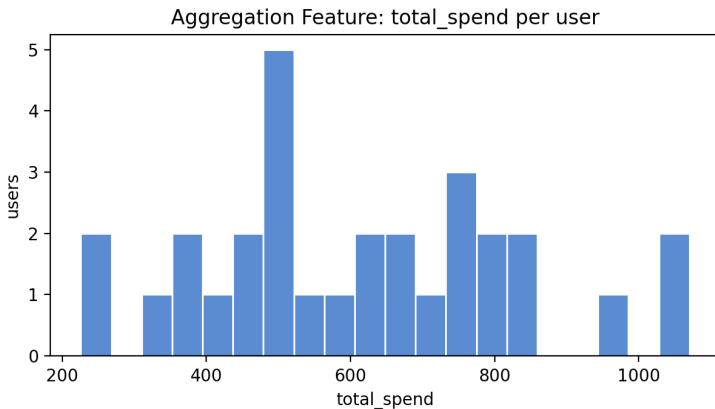
Run from the lecture folder:

```
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

How does cross-validation help detect whether engineered features overfit?