Statistics and Data Analysis

Unit 05 – Lecture 02: Feature Selection Methods (Filter/Wrapper/Embedded)

Tofik Ali

School of Computer Science, UPES Dehradun

February 17, 2026

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

Quick Links

Overview Filter Methods Wrapper/Embedded Exercises Demo Summary

Agenda

Learning Outcomes

- Explain filter methods (variance, correlation, mutual information)

Learning Outcomes

- Explain filter methods (variance, correlation, mutual information)
- Explain wrapper methods (RFE) at a high level

Learning Outcomes

- Explain filter methods (variance, correlation, mutual information)
- Explain wrapper methods (RFE) at a high level
- Explain embedded methods (lasso, tree importance) at a high level

Learning Outcomes

- Explain filter methods (variance, correlation, mutual information)
- Explain wrapper methods (RFE) at a high level
- Explain embedded methods (lasso, tree importance) at a high level
- Discuss pros/cons of each approach

Filter Methods: Key Points

- Fast scoring without training many models

Filter Methods: Key Points

- Fast scoring without training many models
- Examples: variance threshold, correlation with target

Filter Methods: Key Points

- Fast scoring without training many models
- Examples: variance threshold, correlation with target
- May miss interactions

Wrapper/Embedded: Key Points

- Wrapper: search subsets using a model (slow)

Wrapper/Embedded: Key Points

- Wrapper: search subsets using a model (slow)
- Embedded: selection during training (lasso, trees)

Exercise 1: Low variance

If a feature is almost constant, keep it?

Solution 1

- Usually no; low variance adds little information.

Exercise 2: Redundant features

Two features have corr=0.99. What might you do?

Solution 2

- Drop one or use regularization/PCA.

Exercise 3: Wrapper trade-off

Why is RFE slower than filters?

Solution 3

- It trains many models on many subsets.

Mini Demo (Python)

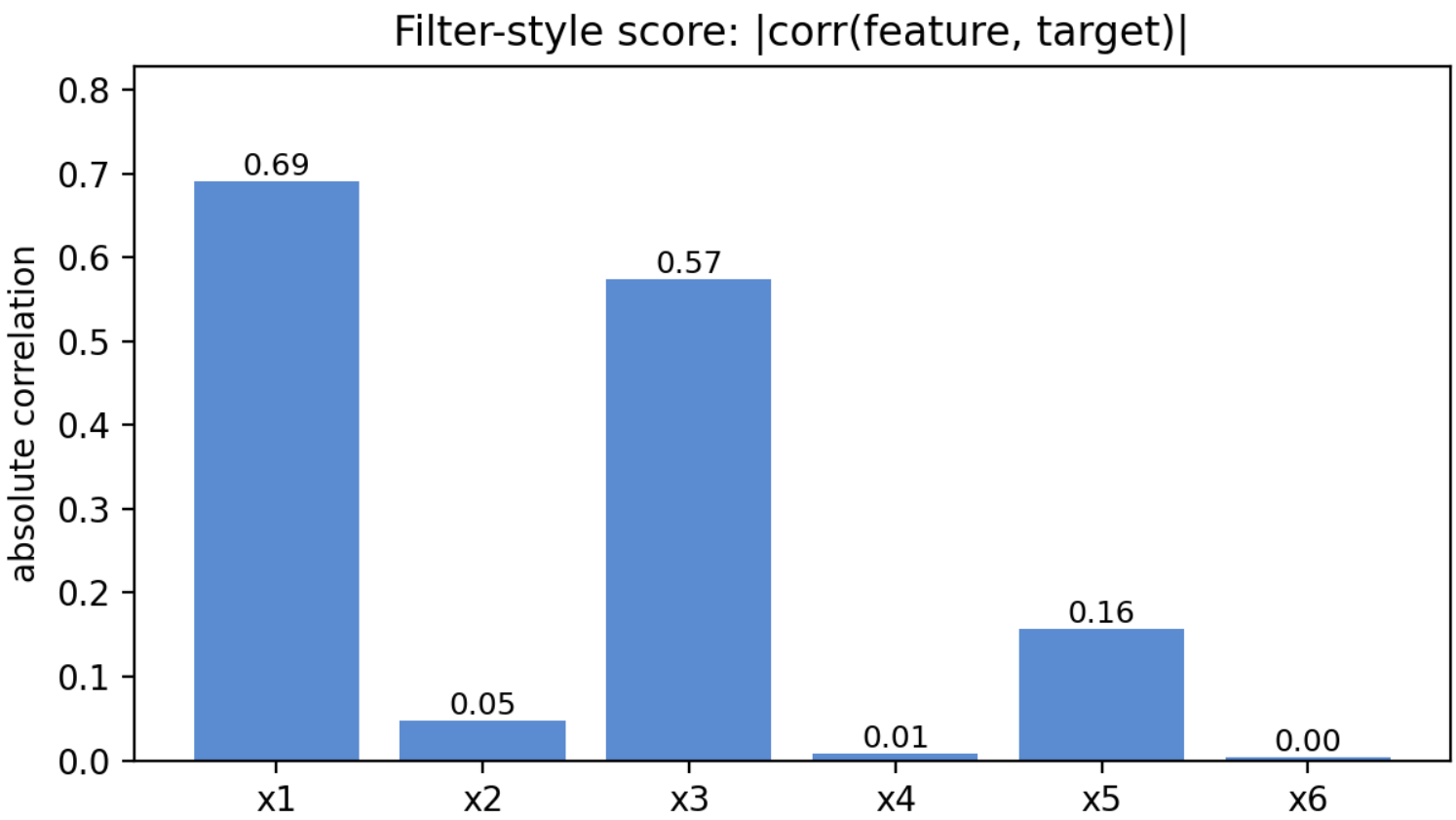
Run from the lecture folder:

```
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

When would you prefer a fast filter method over a wrapper method?