Statistics and Data Analysis

Unit 04 – Lecture 06: VIF, AIC/BIC, Ridge and Lasso (Part 1)

Tofik Ali

School of Computer Science, UPES Dehradun

February 17, 2026

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

Quick Links

Overview VIF Ridge/Lasso Exercises Demo Summary

Agenda

Learning Outcomes

- Compute and interpret VIF (basic)

Learning Outcomes

- Compute and interpret VIF (basic)
- Explain AIC/BIC as model selection criteria (intuition)

Learning Outcomes

- Compute and interpret VIF (basic)
- Explain AIC/BIC as model selection criteria (intuition)
- Write ridge and lasso objectives

Learning Outcomes

- Compute and interpret VIF (basic)
- Explain AIC/BIC as model selection criteria (intuition)
- Write ridge and lasso objectives
- Explain coefficient shrinkage and feature selection idea

VIF: Key Points

- Definition: $\mathrm{VIF}_j = 1/(1 - R_j^2)$

VIF: Key Points

- Definition: $\mathrm{VIF}_j = 1/(1 - R_j^2)$
- Higher VIF -¿ more multicollinearity

VIF: Key Points

- Definition: $\mathrm{VIF}_j = 1/(1 - R_j^2)$
- Higher VIF -¿ more multicollinearity
- Rule of thumb thresholds (5/10)

VIF: Key Formula

$$\mathrm{VIF}_j = \frac{1}{1 - R_j^2}$$

Ridge/Lasso: Key Points

- Ridge uses L2 penalty (shrinks)

Ridge/Lasso: Key Points

- Ridge uses L2 penalty (shrinks)
- Lasso uses L1 penalty (can set some to 0)

Ridge/Lasso: Key Points

- Ridge uses L2 penalty (shrinks)
- Lasso uses L1 penalty (can set some to 0)
- Scale features before regularization

Ridge/Lasso: Key Formula

$$\min \sum (y - \hat{y})^2 + \lambda \sum \beta_j^2 \quad \text{(ridge)}$$

Exercise 1: Compute VIF

If $R_j^2 = 0.9$, compute VIF_j.

Solution 1

- $\mathrm{VIF}_j = 1/(1 - 0.9) = 10$ (high).

Exercise 2: Ridge vs lasso

Which can produce exact zero coefficients?

Solution 2

- Lasso (L1) can set some coefficients to 0.

Exercise 3: AIC/BIC meaning

Lower AIC/BIC means what (conceptually)?

Solution 3

- Better trade-off between fit and complexity (relative).

Mini Demo (Python)

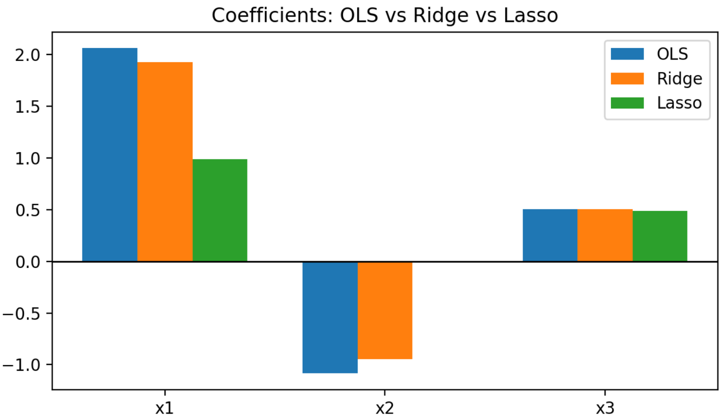
Run from the lecture folder:

```
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

Why can ridge help when predictors are highly correlated?