Statistics and Data Analysis
Unit 04 – Lecture 06 Notes

Tofik Ali

February 14, 2026

Topic

VIF concept; model selection criteria; ridge and lasso regularization (intro).

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

Detailed Notes

These notes are designed to be read alongside the slides. They expand each slide bullet into plain-language explanations, small worked examples, and common pitfalls. When a formula appears, emphasize (1) what each symbol means, (2) the assumptions needed to use it, and (3) how to interpret the final number in the problem context.

VIF

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

Ridge/Lasso

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

Exercises (with Solutions)

Exercise 1: Compute VIF

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

Solution

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

Exercise 2: Ridge vs lasso

Which can produce exact zero coefficients?

Solution

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

Exercise 3: AIC/BIC meaning

Lower AIC/BIC means what (conceptually)?

Solution

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

Exit Question

Why can ridge help when predictors are highly correlated?

Demo (Python)

Run from the lecture folder:

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

Output files:

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

References

- Montgomery, D. C., & Runger, G. C. *Applied Statistics and Probability for Engineers*, Wiley.
- Devore, J. L. *Probability and Statistics for Engineering and the Sciences*, Cengage.
- McKinney, W. *Python for Data Analysis*, O'Reilly.