Statistics and Data Analysis
Unit 04 – Lecture 04 Notes

Tofik Ali

February 14, 2026

Topic

Polynomial regression for curvature; logistic regression for classification; basic evaluation metrics.

Learning Outcomes

- Explain polynomial features for modeling curvature
- Recognize overfitting risk with high degree
- Write logistic regression probability model (sigmoid)
- Compute precision and recall from a confusion matrix

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.

Polynomial Regression

- Add features $x, x^2, x^3, \ldots$
- Still linear in parameters
- Choose degree using validation

Logistic Regression

- Outputs probability in (0,1)
- Threshold converts probability to class label
- Evaluate using confusion matrix / ROC

Exercises (with Solutions)

Exercise 1: Polynomial features

For degree-2 polynomial, what features do we use from x?

Solution

- Use $1, x, x^2$ (intercept + linear + quadratic).

Exercise 2: Precision/recall

TP=30 FP=10 FN=20 TN=40. Compute precision and recall.

Solution

- Precision=30/(30+10)=0.75
- Recall=30/(30+20)=0.60

Exercise 3: Threshold effect

If threshold increases from 0.5 to 0.8, what tends to happen to precision and recall?

Solution

- Precision often increases, recall often decreases.

Exit Question

Why is ROC curve useful when classes are imbalanced?

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.