

# Statistics and Data Analysis

Unit 04 – Lecture 04: Polynomial Regression and Logistic Regression

Tofik Ali

School of Computer Science, UPES Dehradun

February 14, 2026

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

Overview  
o

Polynomial Regression  
o

Logistic Regression  
oo

Exercises  
oooooo

Demo  
oo

Summary  
oo

# Quick Links

Overview

Polynomial Regression

Logistic Regression

Exercises

Demo

Summary

# Agenda

1 Overview

2 Polynomial Regression

3 Logistic Regression

4 Exercises

5 Demo

6 Summary

# Learning Outcomes

- Explain polynomial features for modeling curvature

# Learning Outcomes

- Explain polynomial features for modeling curvature
- Recognize overfitting risk with high degree

# Learning Outcomes

- Explain polynomial features for modeling curvature
- Recognize overfitting risk with high degree
- Write logistic regression probability model (sigmoid)

# 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

# Polynomial Regression: Key Points

- Add features  $x, x^2, x^3, \dots$

# Polynomial Regression: Key Points

- Add features  $x, x^2, x^3, \dots$
- Still linear in parameters

# Polynomial Regression: Key Points

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

# Logistic Regression: Key Points

- Outputs probability in (0,1)

# Logistic Regression: Key Points

- Outputs probability in (0,1)
- Threshold converts probability to class label

# Logistic Regression: Key Points

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

# Logistic Regression: Key Formula

$$P(y = 1 | x) = \frac{1}{1 + e^{-(\beta_0 + \beta^T x)}}$$

# Exercise 1: Polynomial features

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

# Solution 1

- 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 2

- 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 3

- Precision often increases, recall often decreases.

# Mini Demo (Python)

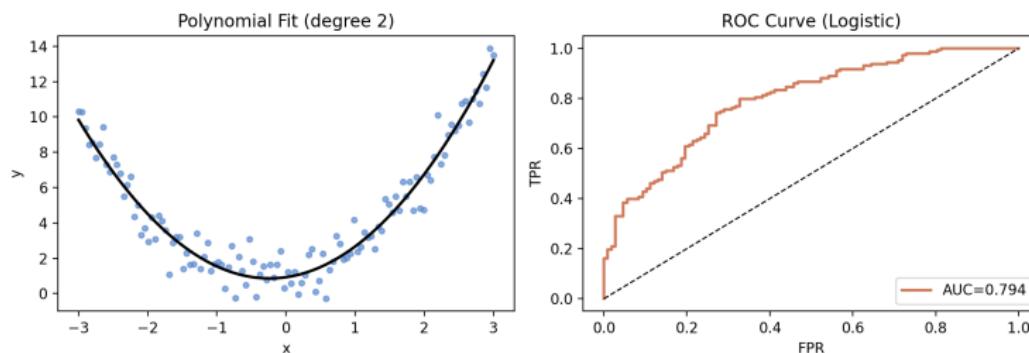
Run from the lecture folder:

```
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 is ROC curve useful when classes are imbalanced?