Lecture 3: Mathematical Foundation and Theory in Pattern Recognition and Anomaly Detection

Dr. Amit Ranjan

School of Computer Science

Test to Geometric Requirement for Curve Algebraic

- **Concept of Curve Fitting**: Curve fitting is the process of finding a mathematical function that best represents a set of data points.
- The function minimizes the difference (error) between the actual data points and the estimated curve.
- It is widely used in machine learning, physics, finance, and engineering.
- Curve fitting is not limited to finding a single function; instead, it involves selecting the best model based on the nature of the dataset. Various techniques exist, including:
  - Linear Regression: Used when the relationship between x and y is approximately a straight line.
  - Polynomial Regression: Used when the relationship follows a nonlinear trend.

Test to Geometric Requirement for Curve Algebraic

Mathematically, given a set of observed data points:

$$D = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$$

We want to find a function $f(x)$ such that:

$$y_i \approx f(x_i) + \epsilon_i, \quad \text{where} \quad \epsilon_i \text{ is the error term}$$

The goal is to minimize the sum of squared errors:

$$S = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

Test to Geometric Requirement for Curve Algebraic

- **Importance of Curve Fitting in Pattern Recognition**:
  - Helps in smoothing noisy data
  - Provides an interpretable model
  - Reduces noise and generalizes data efficiently.
  - Enables pattern recognition by detecting relationships between variables.

Types of Curve Fitting

Least Squares Curve Fitting: A common method of fitting a curve to data is the **Least Squares Method**, which minimizes the squared differences between actual and predicted values.

Least Squares Curve :

For a simple linear model:

$$f(x) = ax + b$$

We calculate the best values of a (slope) and b (intercept) using:

$$a = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{\sum y_i - a \sum x_i}{n}$$

Types of Curve Fitting

Formula for Slope (a):

$$a = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

- n: Number of data points.
- x_i: Values of the independent variable.
- y_i: Values of the dependent variable.

This equation calculates the slope by considering the relationship between x and y while normalizing for the number of data points and spread of the x-values.

Formula for Intercept (b):

$$b = \frac{\sum y_i - a \sum x_i}{n}$$

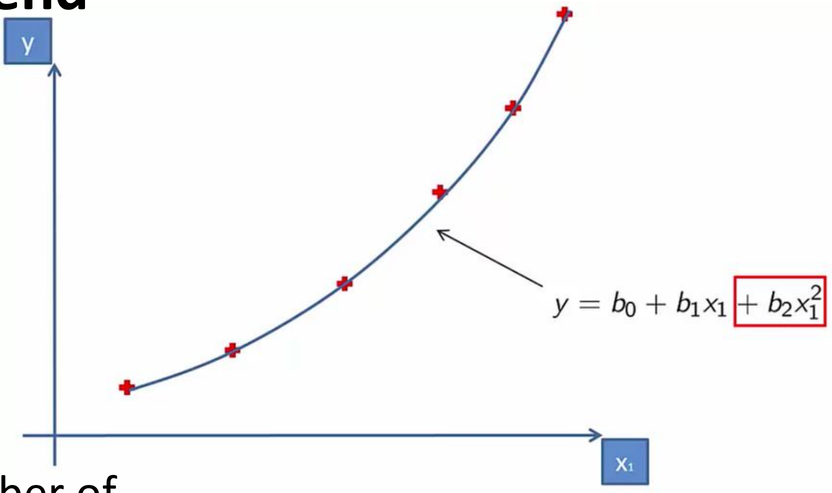
This adjusts the line vertically to best fit the data points, ensuring the overall error is minimized.

Types of Curve Fitting

- **Polynomial Curve Fitting:** Polynomial curve fitting is an extension of linear curve fitting that models nonlinear relationships. The general equation for an n-degree polynomial is:

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

- Polynomial curve fitting is useful when data exhibits a clear **trend but not a linear one**. The **degree of the polynomial** must be chosen carefully:
  - **Low-degree polynomial**: May underfit the data.
  - **High-degree polynomial**: May overfit and introduce unnecessary complexity.

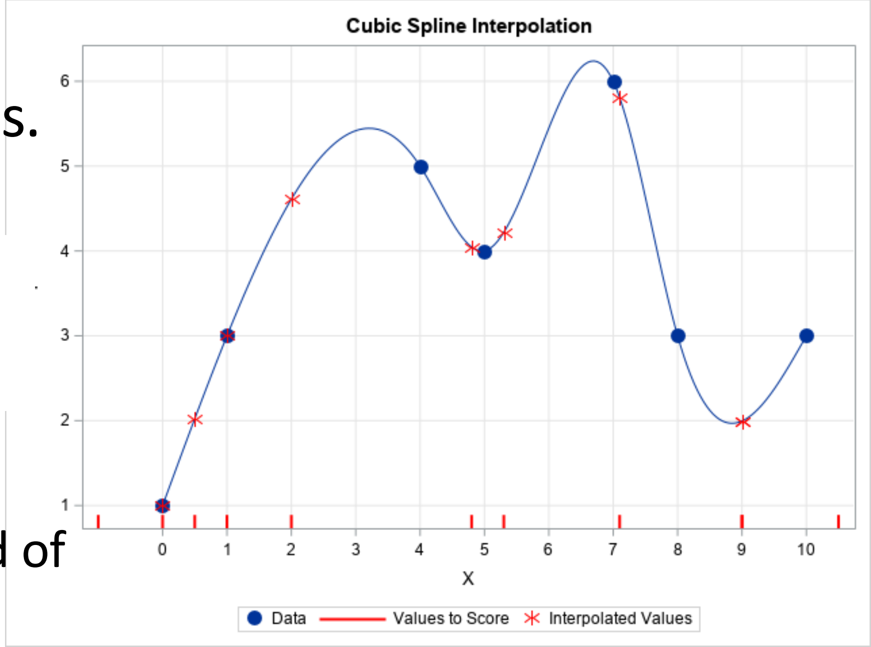


- **Example: Predicting Housing Prices**
  - Housing prices depend on **multiple factors** like square footage, location, number of bedrooms, and economic trends. If a **linear regression** model is insufficient, a **quadratic or cubic polynomial regression** may be more appropriate

Types of Curve Fitting

- **Spline Fitting:** Splines are used when **piecewise smooth** curves are needed. Spline fitting involves piecewise polynomial functions that maintain smoothness at **knots** (joining points).
- Spline fitting is used when the data has **local variations**. The primary advantage of splines over polynomial regression is **better flexibility** and avoiding extreme variations at endpoints. A **cubic spline function** is defined as:

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$$



- **Example: Flight Path Optimization**
  - Air traffic control uses splines to **smooth aircraft trajectories**. Instead of direct straight-line segments, splines allow **gradual turns**, improving fuel efficiency and safety.

Probability Theory Basics

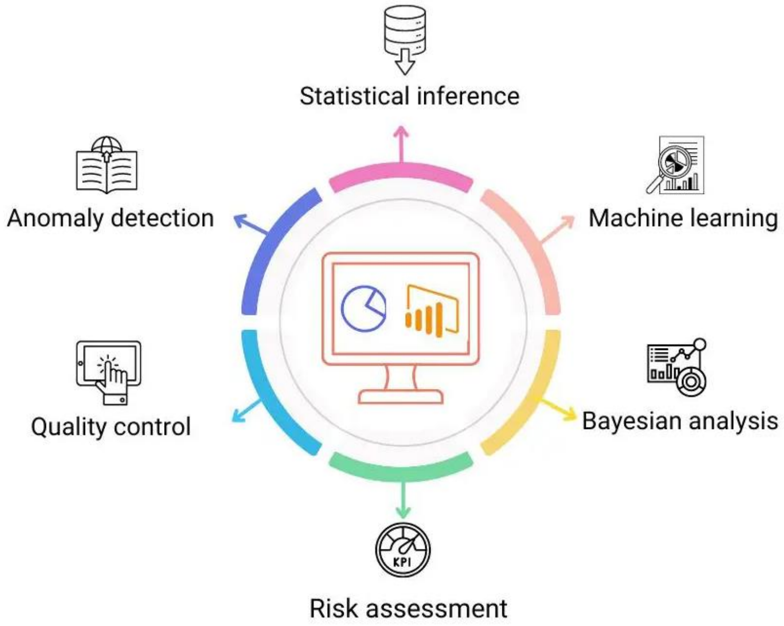
Definition: Probability theory provides a framework to model uncertainties in data.

Probability is the **likelihood of an event occurring**, expressed as:

$$P(E) = \frac{\text{Favorable Outcomes}}{\text{Total Outcomes}}$$

Example: Tossing a fair coin:

$$P(\text{Heads}) = 0.5, \quad P(\text{Tails}) = 0.5$$



Probability Theory Basics

- **Random Variables:** A random variable is a variable whose value is unknown or a function that assigns values to each of an experiment's outcomes.
- The use of random variables is most common in probability and statistics, where they are used to quantify outcomes of random occurrences.
- Risk analysts use random variables to estimate the probability of an adverse event occurring.
- Let's say that the random variable, Z, is the number on the top face of a die when it is rolled once. The possible values for Z will thus be 1, 2, 3, 4, 5, and 6. The probability of each of these values is 1/6 as they are all equally likely to be the value of Z.
- For instance, the probability of getting a 3, or P (Z=3), when a die is thrown is 1/6, and so is the probability of having a 4 or a 2 or any other number on all six faces of a die.

Probability Theory Basics

Discrete Random Variables

- Discrete random variables take on a countable number of distinct values. Consider an experiment where a coin is tossed three times. If X represents the number of times that the coin comes up heads, then X is a discrete random variable that can only have the values 0, 1, 2, or 3 (from no heads in three successive coin tosses to all heads). No other value is possible for X.

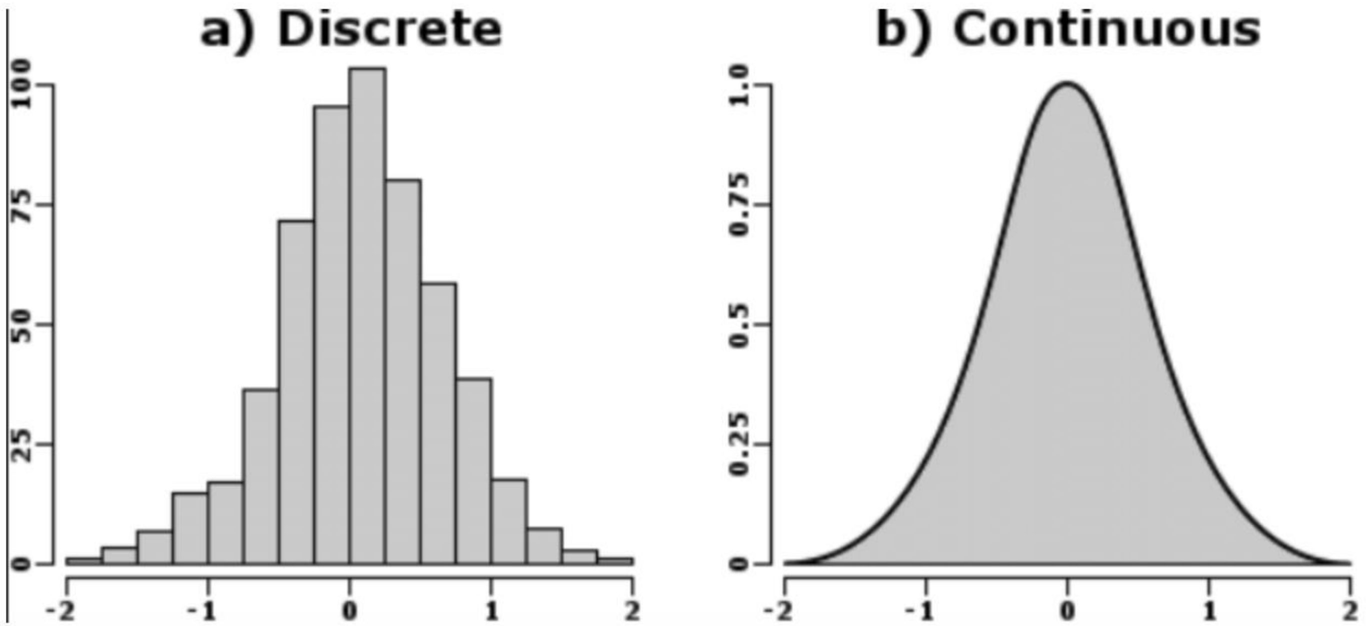
Continuous Random Variables

- Continuous random variables can represent any value within a specified range or interval and can take on an infinite number of possible values. An example of a continuous random variable would be an experiment that involves measuring the amount of rainfall in a city over a year or the average height of a random group of 25 people.

Probability Distributions

Probability Distribution

We know that Probability Distribution is the listing of the probabilities of the events that occurred in random experiments. Probability Distribution is of two types: i) Discrete Probability Distribution and ii) Continuous Probability Distribution.

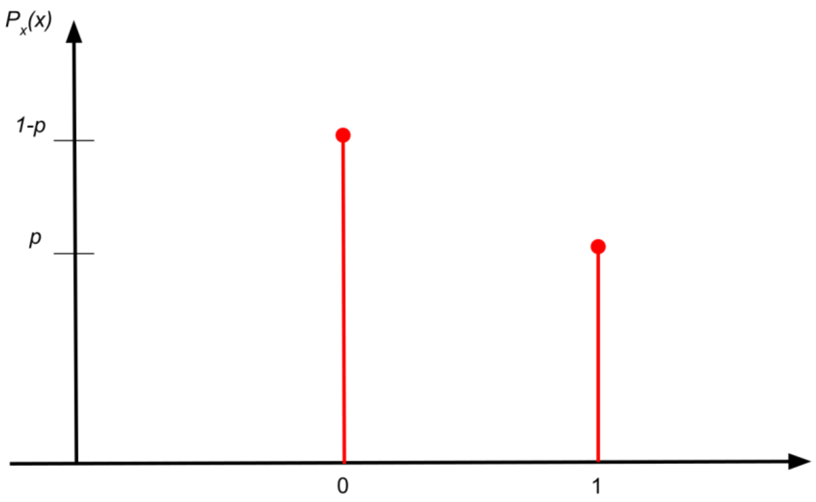


Probability Distributions

Bernoulli Distribution: The Bernoulli Distribution is a discrete distribution where the random variable X takes only two values, 1 and 0, with probabilities p and q where $p + q = 1$. If X is a discrete random variable, then X is said to have a Bernoulli distribution if

$$P(X = x) = \begin{cases} p^x q^{1-x} & for\ x = 0\ and\ 1 \\ 0 & otherwise \end{cases}$$

Here $X = 0$ stands for failure and $X = 1$ for success.



Probability Distributions

Binomial Distribution: The Binomial Distribution is an extension of the Bernoulli Distribution. There are n numbers of finite trails with only two possible outcomes, The Success with probability p and The Failure with q probability. In this distribution, the probability of success or failure does not change from trail to trail, ***i.e., trails are statistically independent***. Thus, mathematically a discrete random variable is said to be Binomial Distribution if,

$$P(X = x) = \begin{cases} {}^{n}_{x}C\, p^{x} q^{n-x} & ;\ x = 0,1,2,\ldots,n \\ 0 & ;\ otherwise \end{cases}$$

Probability Distributions

Poisson's Distribution: The Poisson's Distribution is a limiting case of Binomial Distribution when the trails are infinitely large, i.e., $n \to \infty$, p, the constant probability of success of each trail is indefinitely small, i.e., $p \to 0$ and the mean of Binomial distribution $[np = \lambda]$ is finite.

Mathematically, a random variable X is said to be a Poisson Distribution if it assumes only non-negative values and its PMF is given by

$$p(x, \lambda) = P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad ; \quad x = 0,1,2, \ldots \, and \, \lambda > 0$$

λ is known as parameter of distribution.

Probability Distributions

Normal Distribution: Normal Probability Distribution is one of the most important probability distributions mainly due to two reasons:

a) It is used as a sampling distribution (extracting/creating data samples).
b) It fits many natural phenomena, including human characteristics such as height, weight, etc.

Apart from this, if we talk about the data coming from different sensors, noises present in those signals tend to follow the Gaussian/normal distribution. The Normal Distribution is a continuous probability function that describes how the values of a variable are distributed. It is a systematic distribution where most of the observations cluster around the central peak and the probabilities for values further away from the mean taper off (fade out) equally in both directions.

Mathematically it can be said that, if X is a continuous random variable, then X is said to have normal probability distribution if its PDF is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma}} \; ; -\infty \leq x \leq \infty; -\infty \leq \mu \leq \infty; \sigma > 0$$

Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation (MLE)

MLE finds parameter values that **maximize the probability** of observing the given data.

$$L(\theta) = P(X|\theta)$$

The log-likelihood function is:

$$\log L(\theta) = \sum_{i=1}^{n} \log P(x_i|\theta)$$

Example: Estimating the Probability of a Biased Coin

If we flip a coin 100 times and get 65 heads:

$$L(p) = p^{65}(1-p)^{35}$$

Solving for p:

$$p = \frac{65}{100} = 0.65$$