Machine Learning for Engineering and Science Applications

Introduction to Probability Theory

Discrete and Continuous Random Variables

Introduction

- Probability -- Mathematical framework for representing uncertainty

- Multiple sources of uncertainty in Engineering and Science
  - Inherent randomness in system
    - Example : Quantum mechanics
  - Incomplete data/observability
    - Example: Macroscopic descriptions
  - Incomplete modeling
    - Example : Weather Models

Dual use of probability ideas in Machine Learning

- Constructing Learning systems
  - Incorporate probabilistic algorithms by trying to mimic human reasoning about uncertainty
  - Probabilistic Models
- Analyzing Learning systems
  - Even deterministic learning systems are only correct part of the time. Their output can, therefore, be analyzed probabilistically.
  - Probabilistic analysis of deterministic/probabilistic models

Frequentist vs Bayesian

Statement – 60% chance of rain tomorrow

Two *interpretations* of probability

1. Frequentist
   - Depends on proportion of event in infinite sample space
   - Objective measure
2. Bayesian
   - Measures degree of belief
   - Subjective

- Mathematics of resulting probabilities works the same way
- P(Disease 1) = 0.1, P(Disease 2) = 0.2, P(D 1&D 2) = 0.02 , if they are independent

Definitions

- Random experiment – Experiment that results in different outcomes despite being seemingly similar conditions.
  - Example – Tossing of a coin, throwing of a dice, rainfall amount

- Sample space – Set of all possible outcomes of a random experiment.
  - Example : Tossing of a coin. $S = \{H, T\}$
  - The sample space we choose depends on the purpose of analysis
  - Example : Diameter of a manufactured pipe. S could be

    $S = \mathbb{R}^{+} = \{x \mid x > 0\}$ OR

    $S = \{low, medium, high\}$ OR

    $S = \{satisfactory, unsatisfactory\}$

Random Variables

- Useful to denote outcomes of random experiments by number
- Can be done even for categorical outcomes
- The variable that associates a number with an outcome of a random experiment is called a random variable
- Notation – The random variable is denoted by a capital letter (e.g. X) and its <u>*value*</u> is denoted by a small letter (e.g. x).
  - Example : The rainfall on a particular day is a random variable R.

We can ask "What is the probability that the rainfall is greater than 10mm?" by the mathematical notation $P(R>10) = ?$

Probability Distributions

A probability distribution tells us how likely a random variable is to take each of its possible states.

- Discrete Random Variable (RV)
  - Has finite (or countably infinite) range
  - Example – No. of typographical errors, no. of diagnostic errors, etc
  - Probability measured by Probability Mass Function (PMF)

- Continuous Random Variable (RV)
  - Has real number interval for its range.
  - Example – Temperature, Pressure, Voltage, Height, Current, etc
  - Probability measured by Probability Density Function (PDF)

Probability Mass function

- Discrete Variable -> Probability Mass Function (PMF)
  - PMF -- List of possible values along with their probabilities
  - Example

  X : Number that comes up on throw of a *biased* die

  P(X=1) = 0.1 P(X=1) = 0.1 P(X=3) = 0.2

  P(X=4) = 0.2 P(X=5) = 0.2 P(X=6) = 0.2
- To be a PMF for a random variable X, a function P satisfies:
  - Domain of P is the set of all possible states of X
  - $0 \leq P(X = x) \leq 1$
  - $\sum_{x \in X} P(X = x) = 1$
- Uniform random variable: $P(x = x_i) = \frac{1}{k}$
- Analogous to a point load

Probability Distributions (contd)

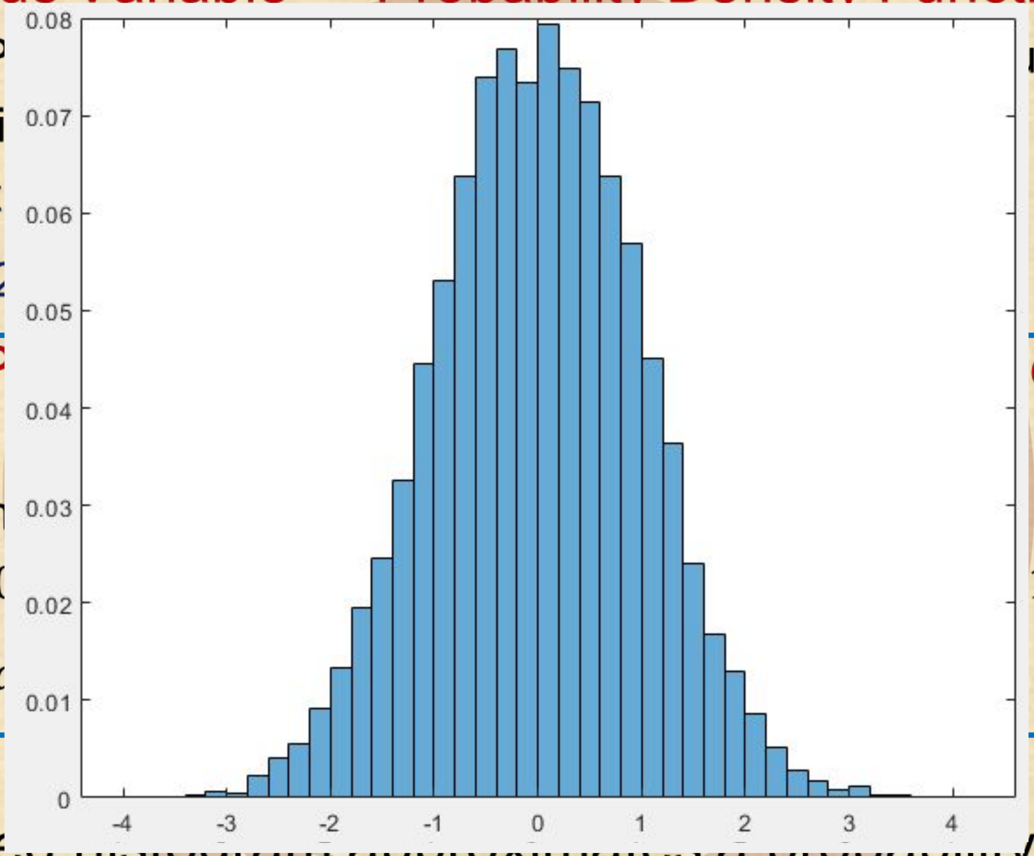
- Continuous Variable -> Probability Density Function (PDF)
  - PDF – P … unit length"
  - Like a di

R : Amount

P(10< R < 2

To be a P … on P satisfies:

- The Don
- $\forall x \in X,$ (… 1
- $\int_X p(x)d$



- Normalized histogram approximates a probability density function