ML + MLOPS

Probability :- Degree of measurement of uncertainity. → Sample Space → Events

Degree of belief :- Absolute certainity → 1 to 0.

Law of Large nos :- more expts. we run, we will tend to get the expected probability.

↳ Gambler's fallacy :- When a certain individual believes which is less likely or more likely.

Random Variable → It has a probability associated with it.

↓ Discrete ↓ Continuous

PMF :- Probabilities of given discrete variable.

PDF :- " " continuous variable.

Expected Value

↳ Mean of probability distribution. $E = \sum x P(x)$

↳ expected probability.

$\bar{y} \approx E(x)$ → law of large nos.

Central Tendency → identifying the central position of the data.

(1) Mean :- Continuous data

↳ most common. * actual value present in data.

↳ If data has outliers, then mean gives wrong interpretation of data.

(2) Median :- Middle score of arranged data

↳ less affected by outliers

(3) Mode :- most freq. score in the dataset

Quantiles

↳ sample divided into equal-sized subgroups. Dividing into areas of equal probability.

Page No.: _____

→ Quartiles :- Divides into 4 equal parts
→ Percentiles :- " " 100 " "
→ Deciles :- " " 10 " ".

* Quartiles : Divides the dataset in 4 equal parts
↳ Q_1, Q_2, Q_3 quartiles.

2 | 4 | 4 | 5 | 6 | 7 8

Q_1 (lower quartile) | Q_2 (middle quartile) | Q_3 (Upper quartile)

central point b/w
Q_1 ⇒ Smallest and median

$Q_1 = n \times 1/4$

Q_3 ⇒ middle value b/w Q_2 & highest score

Interquartile Range
↳ $Q_3 - Q_1$
↳ Resistant of Outliers

Percentile :- Certain % of scores fall below that number.

Decile :- 10% ⇒ 1st decile
60% ⇒ 6th decile

Variance :- How far data points differ from mean.

↑ variance ↑ is called

$\sigma^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1}$

σ :- std. deviation.

↳ value of observation

Covariance :- Relationship b/w two random variables (directional)

↳ if both increase ⇒ +ve covariance

↳ if ~~one decrease~~, & increase ⇒ -ve covariance.

Date: Page No.:

$$Cov(X,Y) = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Correlation :- defines the strength of relationship. (Ranges from -1 to 1)

↳ Pearson coefficient : $P_{xy} = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$ σ_x :- Standard deviation of x, σ_y :- Standard deviation of y.

Joint Probability Distribution :- two probabilities are at same time.

↳ Conditional probability : $P(A|B) = \frac{P(A \cap B)}{P(B)}$

If A & B are disjoint

$A \cap B = \phi \Rightarrow P(A|B) = 0$

Distributions :- ① Uniform distribn :- Probabilities is constant over a range.

$$f(x) = \frac{1}{b-a} \begin{cases} a \le x \le b \\ 0, \text{ otherwise} \end{cases}$$

② **Normal Distribution :-** Data which is closer is more freq. to occur.

↳ if graph is normally distributed

→ mean = median = mode

→ Symmetric about curve.

Empirical Rule :- Where the most of values in ND.

↳ $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

68% lies in 1SD
95% lies in 2SD
99.7% lies in 3SD

↳ $z = \frac{x-\mu}{\sigma}$

Geometric Distribn :- $(1-p)^{n-1} p$

↳ $\mu = \frac{1}{p}$; $\sigma^2 = \frac{1-p}{p^2}$, $\sigma = \sqrt{\frac{1-p}{p^2}}$