DATE
PAGE

Statistics - 6

(1) (d) All of the above

(2) (a) discrete

(3) (a) Pdf

(4) (c) mean

(5) (a) variance.

(6) (a) variance.

(7) (c) 0 and 1

(8) (b) bootstrap

(9) (b) summarized.

(10). Both histogram and boxplots are used to explore and present the data in an easy and understandable manner.

. Histograms are preffered to determine the underlying probability distributions

of a data.

- Box Plots on the other hand are more useful when comparing between several data sets.
They are less detailed than histograms and take up less space.

- Although histograms are better in distribution of data, you can use a box plot to tell if distribution is symmetric or skewed.

(11) • The evaluation metrics used in regression model (continuous output) & a classification model (numeric or binary output) are different.

- In classification problems, we have 2 type of algorithms:

→ a. Class output: ~~Algo~~

Algorithms like KNN & SVM create a class output.

→ Probability output:

Algorithms like Logistic regression, random forest, gradient boosting, Adaboost give

probability results

- In regression models, we do not have such inconsistencies in output. The output is always continuous in nature & requires no further treatment

---

(12) • Statistical significance can be accessed using hypothesis testing :-

- Stating a null hypothesis which is usually the opposite of what we wish to test
- Then, chose a suitable statistical test & statistics used to reject the null hypothesis
- Also, chose a critical region for the statistics to lie in that is extreme enough for null hypothesis to be rejected.
- Calculate the observed test statistics from data and check whether it lies in critical region

(13) Examples of data that does not have gaussian distribution are :-

- Weibul distribution : life data such as survival times of a product
- Exponential distribution : growth data such as bacterial growth
- Poisson distribution : rare events such as number of accidents
- Binomial distribution :- Proportion data such as percent defects

---

(14)
- Unlike the mean the median value does not depend on all the values of dataset.
- In the presence of outliers, effect on the median is smaller
- When you have skewed distribution median is better measure than mean.

DATE
PAGE

(15) - Likelihood measures the goodness of fit of a statistical model to a sample ~~data~~ of data for a given value of unknown parameter.

- It is formed from joint probability distribution of the sample, but viewed and used as a function of the parameter only thus treating random variables as 'fixed at the observed values.