SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

Re-accredited by NAAC with A+ grade, Accredited by NBA(CSE, IT, ECE, EEE & Mechanical)

Approvedy by AICTE, New Delhi, Recognized by UGC, Affiliated to Anna University, Chennai

Department of MCA

Topic: **Prediction Error**

Course

19CAP704
Big Data Analytics

Unit I

Introduction to Big data

Elective

III Semester / II MCA

Problem

A company manufacturing disposable tableware such as paper plates forecasts of each of hundreds of items every month.

Time series data showed a range of patterns, some with trends, some seasonal etc..

Software produced forecasts that did not seem sensible..

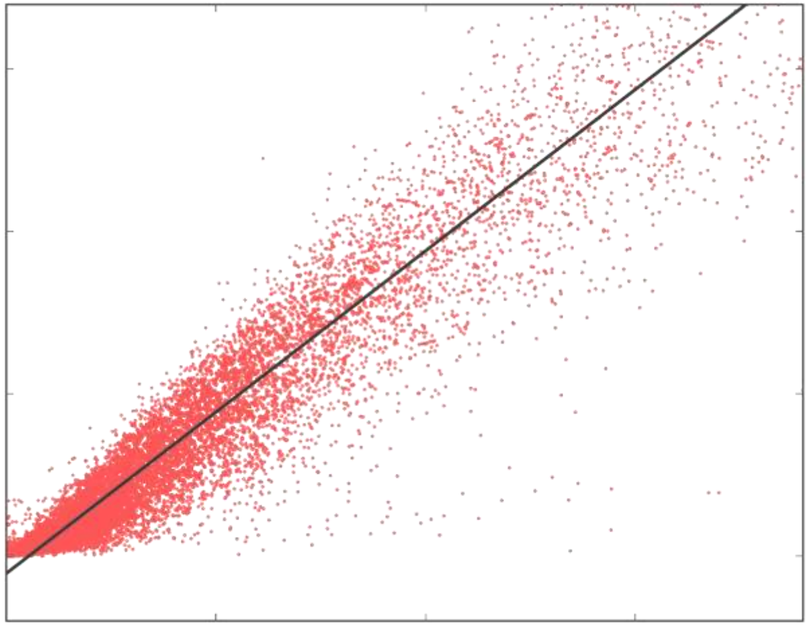
Predictive Analysis

- Practice of extracting insights from data set with help of data mining, statistical modeling and machine learning techniques and using it to predict unobserved or unknown event
- Identifying cause – effect relationship across variables
- Applying observed patterns to unknown
- Methods: Regression, classification, time series forecasting, association rule mining, clustering

Regression

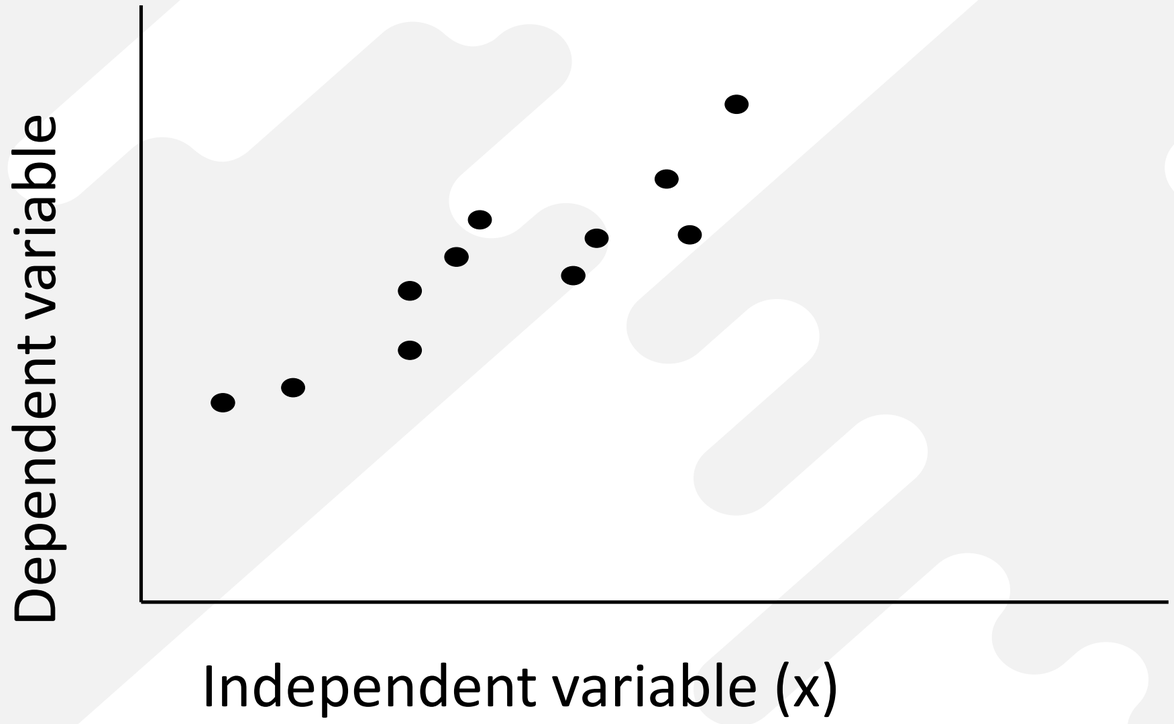
- Determines statistical relationship between two or more variables where a change in a dependent variable is associated with, and depends on, a change in one or more independent variables

- Regression line estimates the average value of y for each x.



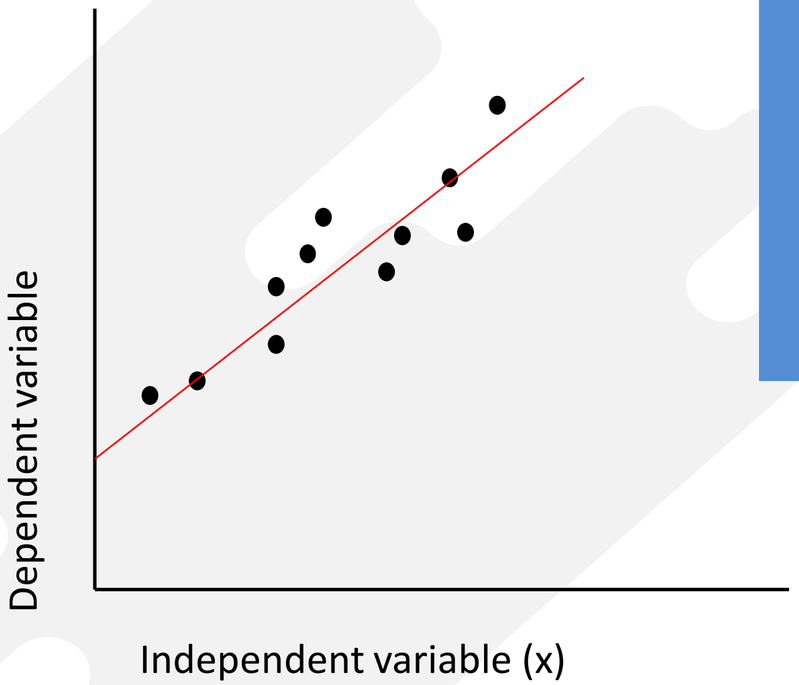
Regression

CAUSE → **EFFECT**



Regression

- The function will make a prediction for each observed data point
- Simple regression fits a straight line to the data.
- ***Regression analysis*** is a statistical process for estimating the relationships among variables

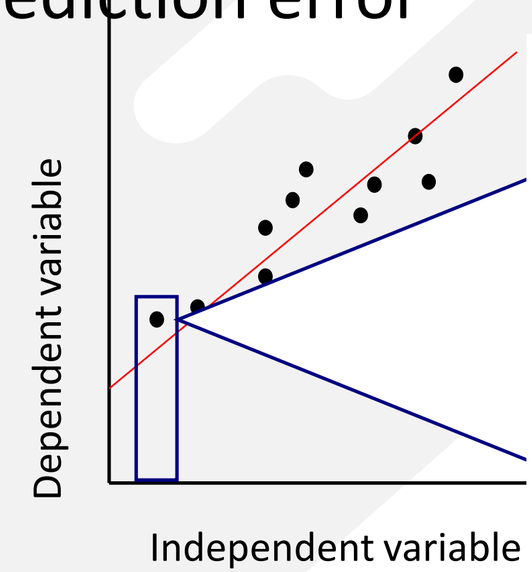


Prediction Error

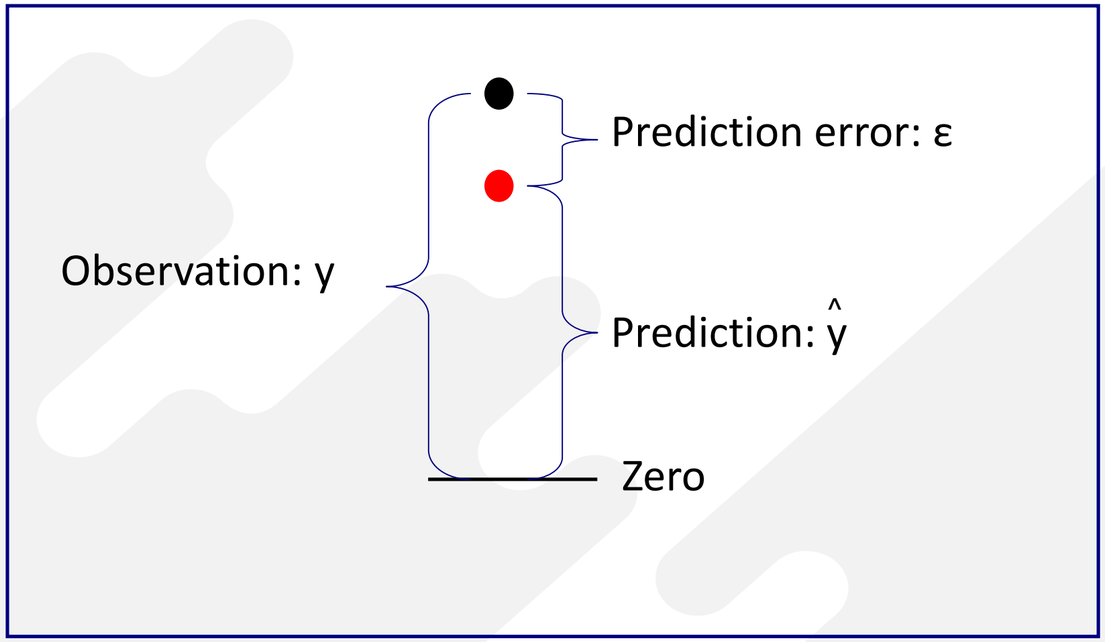
- Parameter estimation technique
- Failure of some expected event to occur
- Difference between the expected value and true value of Y is called prediction error
- Inescapable element of predictive analytics that should also be quantified and presented along with any model, often in the form of a confidence interval that indicates how accurate its predictions are expected to be

Prediction Error

- Failure of some expected event to occur
- Difference between the expected value and true value of Y is called prediction error



Prediction Error

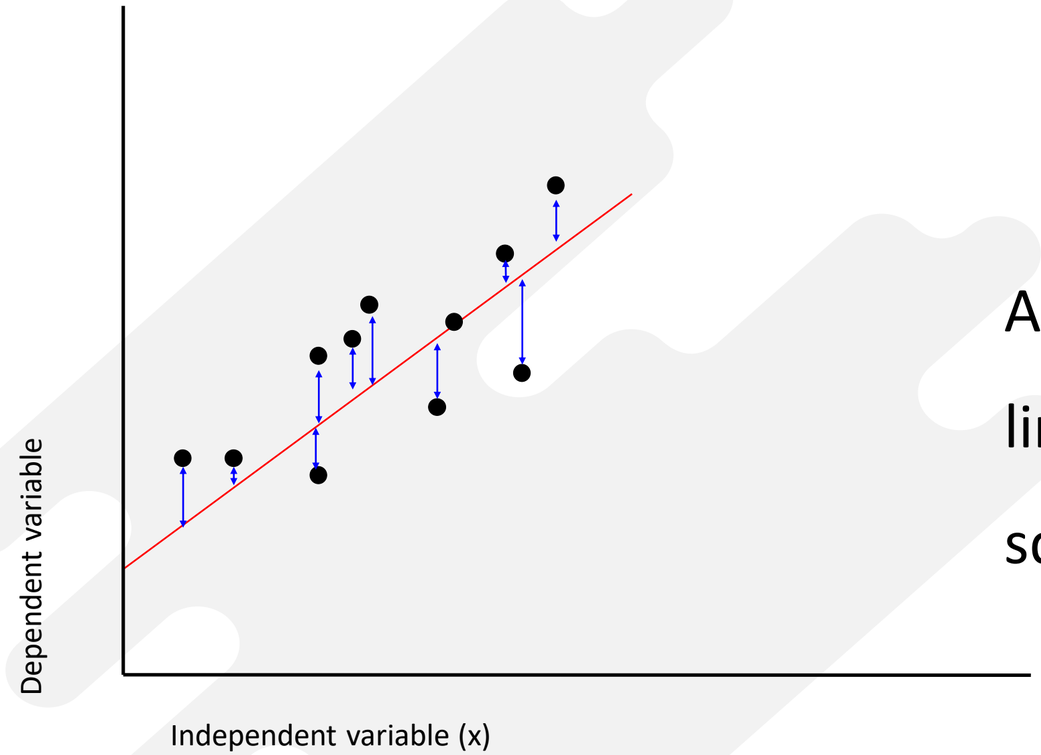


For each observation, the variation can be described as:

$$y = y + \varepsilon$$

Actual = Explained + Error

Prediction Error



A least squares regression selects the line with the lowest total sum of squared prediction errors

Assessment

Assessment

Assessment

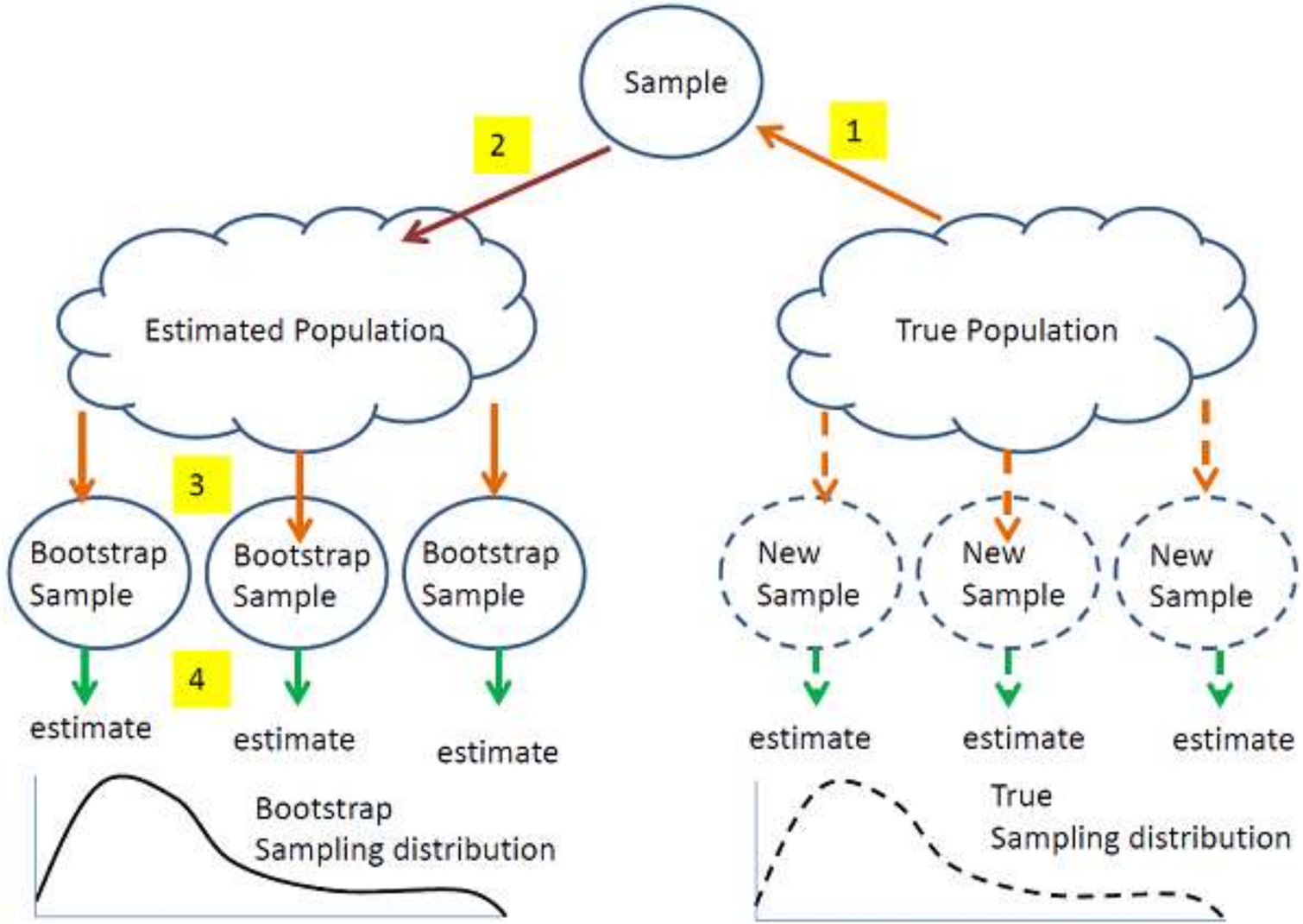
Methods for Prediction Error

- Cross- validation (CV)
  - Estimate the prediction error for data model, when data is limited
  - Train the model with subset of data and test it on the remaining data
    - Repeat this with different subset of data
  - Idea is split the training data into two subsets –
    - One subset is used to train the prediction rule
    - The other subset is used to assess prediction error
  - In machine learning, CV assesses prediction error and trains the prediction rule.

Methods for Prediction Error

- Bootstraps
  - Method of sample reuse that is much more general than cross-validation
  - Idea is to use the observed sample to estimate the population distribution
    - Nonparametric (resampling)
    - Semiparametric (adding noise)
    - Parametric (simulation)

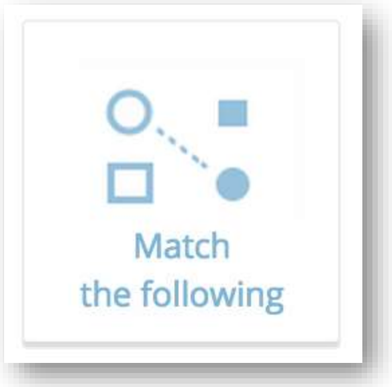
Methods for Prediction Error



Assessment

1. Fitting the model in the statistics
2. Developed data on prediction
3. Data used to assess the prediction rule
4. Bootstrapping
5. Cross Validation

A. Estimate sampling distribution
B. Training the prediction rule
C. Train and test with different subset of data
D. Validation data
E. Training data



References

www.statisticshowto.com/prediction-error-definition/

https://web.stanford.edu/class/msande226/lecture5_prediction.pdf

https://newonlinecourses.science.psu.edu/stat555/node/116/

https://online.stat.psu.edu/stat555/node/116/


thank
you!