Contents

Data science

- descriptive: **what** happened?
- diagnostic: **why** did it happen?
- predictive: what **will** happen?
- prescriptive: what shold I do?

Data driven

data —-> insight —> action

predictive modeling

- classfication
- regression
- forecasting?

descriptive modeling

- cluster analysis/segmenttatioin

discovering patterns and rules

- association

deviation detection

- outlier analysis

Flow Chart

Business understanding

Data understanding

Data prepration

- get the raw data
- preprocessing
  - clearning & filtering
    - outlier

missing data time vs accuracy tradeoffs

- leave as NULL

            - delete rows
            - delete feature with high missingness
            - imputation
            - mean
            - model
    - variable transformation
        * featuring engineering
            - combine attributes: rates, ratios
            - scaling data
            - z-score -> mean 0, std=1
            - scale -> [0, 1]
            - log
            - discretize data
    - variable selection
        * step wise
        * correlation
        * feature importance

Modeling

Evaluation

- signficance tests

Deployment

- training
- predicting