Data Science
Lab Exercise (Naïve Bayes)
Prepared By
Dr Muhammad Atif Tahir

- Create text file for the following data and load data through Pandas libraries
- Build your own Naïve Bayes classifier model using steps below
- Class: $P(C) = N_c/N$
- e.g., $P(No) = 7/10$,
  $P(Yes) = 3/10$
- For discrete attributes:

  $P(A_i | C_k) = |A_{ik}| / N_{ck}$

  where $|A_{ik}|$ is number of instances having attribute A_i and belongs to class C_k
  Examples:
  $P(Status=Married|No) = 4/7$
  $P(Refund=Yes|Yes)=0$

- Normal distribution:
  One for each (A_i, c_i) pair

  $$P(A_i | c_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(A_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

  For (Income, Class=No):
  If Class=No
  sample mean = 110
  sample variance = 2975

- Once Trained, test your model for the cases below

  X1 = {Refund = Yes, Status = Divorced, Income = 90K, Evade = ?}
  X2 = {Refund = No, Status = Married, Income = 60K, Evade = ?}

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

categorical   _categorical_   _continuous_   _class_