Naïve Bayes Classifier

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST

icmoon@kaist.ac.kr

Dataset for Optimal Classifier Learning with Conditional Independent Assumption

Sky	Temp	Humid	Wind	Water	Forecst	EnjoySpt
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

- Previously, $f^*(x) = argmax_{Y=y} P(X = x | Y = y) P(Y = y)$
  - **$P(X=x|Y=y)$ has $(2^d-1)k$ cases**
- Let's apply the conditional independent assumption to the all features of X (=all variables in the vector of x)
- Now, $f^*(x) = argmax_{Y=y} P(X = x | Y = y) P(Y = y)$

$$\approx argmax_{Y=y} P(Y = y) \prod_{1 \leq i \leq d} P(X_i = x_i | Y = y)$$

  - How many parameters after adopting the assumption?
  - **$P(X_i = x_i | Y = y)$ has $(2-1)dk$ cases**
- *You: Wait! The passing-by statistician! Is that right????!!!!*

Naïve Bayes Classifier

- Statistician: Yeah. I know that the assumption is naïve. Why don't you call it as naïve Bayes classifier?

- Given:
  - Class Prior **P(Y)**
  - d conditionally independent features X given the class Y
  - For each X_i, we have the likelihood of $P(X_i|Y)$
- **Naïve Bayes Classifier Function**
  - $f_{NB}(x) = argmax_{Y=y} P(Y = y) \prod_{1 \leq i \leq d} P(X_i = x_i|Y = y)$

- Naïve Bayes classifier is the optimal classifier
  - If the conditional independent assumptions on X hold
  - If the prior is right
- Any problems????

Problem of Naïve Bayes Classifier

- Problem 1: Naïve assumption
  - Many, many, many cases, the variables of X are correlated
  - Why?
  - Multi-collinearity
- Problem 2: Incorrect Probability Estimations
  - Billionaire
    - Head, Head, Head…
  - MLE with insufficient data
    - There is no chance of Tail!
    - $P(Y=tail) = 0$
  - MAP with stupid prior
    - Is either our dataset or knowledge good enough to estimate the prior?
- Problem 2 is always there!
- Problem 1 is introduced by our assumption!