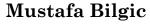
CS 584 - MACHINE LEARNING

TOPIC: NAÏVE BAYES





♦ http://www.cs.iit.edu/~mbilgic



https://twitter.com/bilgicm

CLASSIFICATION

- Input: $\vec{X} = \langle X_1, X_2, \dots, X_n \rangle$
- Output: Y

- We have seen
 - Candidate elimination to find the full version space
 - Decision trees

BAYES CLASSIFIER

$$P(Y \mid \vec{X}) = \frac{P(\vec{X} \mid Y)P(Y)}{P(\vec{X})} = \frac{P(Y)P(X_1, X_2, ..., X_n \mid Y)}{P(X_1, X_2, ..., X_n)}$$

$$P(X_1, X_2, ..., X_n) = \sum_{y} P(Y = y) P(X_1, X_2, ..., X_n \mid Y = y)$$

Assuming all variables are binary, how many independent parameters are needed for the Bayes classifier?

Naïve Bayes Assumption

$$X_i \perp X_j \mid Y$$

Naïve Bayes

Bayes rule:

$$P(Y \mid X_1, X_2, ..., X_n) = \frac{P(Y)P(X_1, X_2, ..., X_n \mid Y)}{\sum_{y} P(y)P(X_1, X_2, ..., X_n \mid y)}$$

Assuming $X_i \perp X_j \mid Y$, naïve Bayes:

$$P(Y \mid X_1, X_2, ..., X_n) = \frac{P(Y) \prod P(X_i \mid Y)}{\sum_{y} P(y) \prod P(X_i \mid y)}$$

Assuming all variables are binary, how many independent parameters are needed for the naive Bayes classifier?

Naïve Bayes Implementations

- o Bernoulli / categorical naïve Bayes
 - Features are assumed to be binary / categorical
- Multinomial naïve Bayes
 - $P(\vec{X} \mid y)$ is a multinomial distribution
- o Gaussian naïve Bayes
 - Each $p(x_i \mid y)$ is a Gaussian distribution

PARAMETER ESTIMATION

- Given a dataset $\mathcal{D} = \{\langle \vec{X}[m], Y[m] \rangle\}$, how can we estimate
 - P(Y)
 - $P(X_i \mid Y)$
- Intuitive idea: count and normalize
 - But, why is this the right idea? Or, is it even the right idea?

TOPIC SWITCH

• Probability estimation from data

To be continued...