

CS 581 – ADVANCED ARTIFICIAL INTELLIGENCE

TOPIC: NAÏVE BAYES



Mustafa Bilgic



<http://www.cs.iit.edu/~mbilgic>



<https://twitter.com/bilgicm>

BAYES CLASSIFIER

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

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

$$P(X_1, X_2, \dots, X_n) = \sum_y P(Y = y)P(X_1, X_2, \dots, X_n | Y = y)$$

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

EXCURSION

- Maximum likelihood estimation

NAÏVE BAYES ASSUMPTION

$$X_i \perp X_j \mid Y$$

NAÏVE BAYES

Bayes rule:

$$P(Y | X_1, X_2, \dots, X_n) = \frac{P(Y)P(X_1, X_2, \dots, X_n | Y)}{\sum_y P(y)P(X_1, X_2, \dots, X_n | y)}$$

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

$$P(Y | X_1, X_2, \dots, X_n) = \frac{P(Y) \prod P(X_i | Y)}{\sum_y P(y) \prod P(X_i | y)}$$

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

EXAMPLE

- See OneNote

NAÏVE BAYES IMPLEMENTATIONS

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

READING

- <http://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf>
- https://en.wikipedia.org/wiki/Naive_Bayes_classifier
- https://scikit-learn.org/stable/modules/naive_bayes.html

ZERO PROBABILITIES

- Assume feature X_i is T for a particular object.
Further,
 - Assume $P(X_i = T|yes) = 0$ and $P(X_i = T|no) > 0$ and $P(X_j | yes) > 0$ and $P(X_j | no) > 0$ for all other features
 - What is $P(yes | \vec{X})$?
 - Assume $P(X_i = T|yes) = 0$ and $P(X_i = T|no) = 0$ and $P(X_j | yes) > 0$ and $P(X_j | no) > 0$ for all other features
 - What is $P(yes | \vec{X})$?
- One solution: use Bayesian estimation

EXCURSION

- Bayesian estimation

MULTIPLYING SEVERAL PROBABILITY NUMBERS

- Assume we have 10,000 features
- What is $0.9^{10,000}$ using a computer?
- Try `math.pow(0.9, 10000)` in Python
- In Naïve Bayes,
 - $a = P(Y = T) \prod P(X_i|Y = T)$
 - $b = P(Y = F) \prod P(X_i|Y = F)$
 - $P(Y = T|\vec{X}) = \frac{a}{a+b}$
 - If $a = b = 0$ in your code, then what?