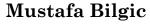
CS578 – Interactive and Transparent Machine Learning

TOPIC: NAÏVE BAYES





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



https://twitter.com/bilgicm

BAYES RULE

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

BAYES CLASSIFIER

- Input: X
- Output: *Y*
- Bayes classifier
 - $P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$

BAYES CLASSIFIER

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

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

NAÏVE BAYES ASSUMPTION

$$X_i \perp X_j \mid Y$$

Naïve Bayes Classifier

$$P(Y|X_{1}, X_{2}, ..., X_{n}) = \frac{P(X_{1}, X_{2}, ..., X_{n}|Y)P(Y)}{P(X_{1}, X_{2}, ..., X_{n})}$$

$$= \frac{P(X_{1}|Y)P(X_{2}|Y) ... P(X_{n}|Y)P(Y)}{P(X_{1}, X_{2}, ..., X_{n})}$$

$$= \frac{P(Y) \prod_{i=1}^{n} P(X_{i}|Y)}{P(X_{1}, X_{2}, ..., X_{n})}$$

SCIKIT-LEARN

- https://scikit-learn.org/stable/modules/naive_bayes.html
- 1. Bernoulli naïve Bayes
- 2. Gaussian naïve Bayes
- 3. Multinomial naïve Bayes
- 4. Complement naïve Bayes

1. Bernoulli Naïve Bayes

- Binary input; k classes; M data points
- Estimate
 - $P(X_i|Y)$ for each binary feature X_i

•
$$P(X_i = True | Y = y_j) = \frac{\#(X_i = True, Y = y_j) + \alpha}{\#(Y = y_j) + 2\alpha}$$

•
$$P(X_i = False | Y = y_i) = 1 - P(X_i = True | Y = y_i)$$

P(Y)

$$P(Y = y_j) = \frac{\#(Y = y_j) + \alpha}{M + k\alpha}$$

- $P(Y|\vec{X}) \propto P(Y) \prod_{i} P(X_i|Y)$
 - Calculate $P(Y) \prod_i P(X_i|Y)$ for each y_i and then normalize

Bernoulli Naïve Bayes Code

 https://scikitlearn.org/stable/modules/generated/sklearn.naive_bayes.BernoulliNB.html#sklearn.n aive_bayes.BernoulliNB

Attributes

- class_count_
- feature_count_
- feature_log_prob_
- class_log_prior_

EVIDENCE COMPUTATION FOR BINARY CLASSIFICATION

• [The lecture]

2. Gaussian Naïve Bayes

- Continuous input; k classes; M data points
- Estimate
 - $p(X_i|Y)$ for each continuous feature X_i
 - $p(X_i|Y=y_i)$ is assumed to be a Gaussian distribution
 - ullet Note that each feature will be represented with a different Gaussian distribution per class
 - $p(X_i|Y=y_j)\sim N(\mu_j^i;\sigma_j^i)$ k different distributions for X_i
 - \bullet P(Y)

$$P(Y = y_j) = \frac{\#(Y = y_j) + \alpha}{M + k\alpha}$$

GAUSSIAN NAÏVE BAYES CODE

<u>https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html#sklearn.naive_bayes.GaussianNB</u>

Attributes

- class_count_
- class_prior_
- theta_
- sigma_

3. Multinomial Naïve Bayes

- Non-negative input; k classes; M data points
- Estimate
 - k multinomial distributions
 - o $P(\vec{X}|Y=y_j)=[p_j^1,p_j^2,...p_j^N]$ -- a multinomial distribution per class
 - P(Y)

$$P(Y = y_j) = \frac{\#(Y = y_j) + \alpha}{M + k\alpha}$$

- ∘ $P(Y|\vec{X}) \propto P(Y) \prod_i P(X_i|Y)$ for all non-zero X_i
- Note the similarities and differences between BernoulliNB and MultinomalNB
- Works well for text classification

Multinomial Naïve Bayes Code

 https://scikitlearn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html#sklear n.naive_bayes.MultinomialNB

Attributes

- class_count_
- feature_count_
- feature_log_prob_
- class_log_prior_

EVIDENCE COMPUTATION FOR TEXT CLASSIFICATION

• [The lecture]