PS5841

Data Science in Finance & Insurance

Naïve Bayes

Yubo Wang

Spring 2022

Columbia University

Bayes Rule / Theorem

$A_1, \ldots, A_n$ form a partition of the entire probability space S.

$$\Pr[A_j | B] = \frac{\Pr[A_j \cap B]}{\mathrm{P}[B]} = \frac{\Pr[B | A_j] \Pr[A_j]}{\sum_{j=1}^{n} \Pr[B | A_j] \Pr[A_j]}$$

COLUMBIA UNIVERSITY

Conditional Probability

- Conditional Probability

$$\Pr[B|A_j] = \frac{\Pr[B \cap A_j]}{\Pr[A_j]}$$

$$\rightarrow \Pr[A_j \cap B] = \Pr[B|A_j]\Pr[A_j]$$

- Law of Total Probability

$$\Pr[B] = \sum_{j=1}^{n} \Pr[B \cap A_j] = \sum_{j=1}^{n} \Pr[B|A_j]\Pr[A_j]$$

Bayes Classifier (1)

- Classifies an observation $\boldsymbol{x}$ to the class with the greatest posterior probability

$$p_k(\boldsymbol{x}) = \Pr(Y = k | X = \boldsymbol{x})$$

- Has the lowest possible error rate out of all classifiers

$$p_k(\boldsymbol{x}) = \frac{\Pr(X = \boldsymbol{x} | Y = k)\Pr(Y = k)}{\sum_{l=1}^{K} \Pr(X = \boldsymbol{x} | Y = l)\Pr(Y = l)}$$

$$= \frac{f_k(\boldsymbol{x})\pi_k}{\sum_{l=1}^{K} f_l(\boldsymbol{x})\pi_l} = \frac{\pi_k f_k(\boldsymbol{x})}{\sum_{l=1}^{K} \pi_l f_l(\boldsymbol{x})}$$

$A_l = (Y = l), B = (X = \boldsymbol{x})$

COLUMBIA UNIVERSITY

Bayes Classifier (2)

$$p_k(\boldsymbol{x}) = \frac{\pi_k f_k(\boldsymbol{x})}{\sum_{l=1}^{K} \pi_l f_l(\boldsymbol{x})}$$

- Estimation
  - prior

$$\hat{\pi}_k = \frac{n_k}{N}$$

  - likelihood

$$\hat{f}_k(\boldsymbol{x}) = ?$$

Columbia University

Naïve Bayes Classifier (1)

- Naïve Assumption: Features are independent of each other

$$f_k(\boldsymbol{x}) = \prod_{j=1}^{p} f_{kj}(x_j), \qquad \boldsymbol{x} = \left(x_1, \dots, x_p\right)^T$$

- Classifies an observation $\boldsymbol{x}$ to the class with the greatest posterior probability

$$p_k(\boldsymbol{x}) = \frac{\pi_k \prod_{j=1}^{p} f_{kj}(x_j)}{\sum_{l=1}^{K} \pi_l \prod_{j=1}^{p} f_{lj}(x_j)}$$

COLUMBIA UNIVERSITY

Naïve Bayes Classifier (2)

$$\Pr(Y = k | X = \boldsymbol{x}) \propto$$
$$\Pr(Y = k) \Pr(X_1 = x_1 | Y = k) \ldots \Pr(X_p = x_p | Y = k)$$

$$p_k(\boldsymbol{x}) \propto \pi_k f_{k1}(x_1) \ldots f_{kp}(x_p)$$

COLUMBIA UNIVERSITY

Naïve Bayes Classifier (3)

Ex: 2 quantitative features & 2 classes

- For a particular observation (x_1, x_2)

$$p_0[(x_1, x_2)]$$
$$= \frac{\pi_0 f_{01}(x_1) f_{02}(x_2)}{\pi_0 f_{01}(x_1) f_{02}(x_2) + \pi_1 f_{11}(x_1) f_{12}(x_2)}$$
$$= const \cdot \pi_0 f_{01}(x_1) f_{02}(x_2)$$

- Classify the observation to class 0 if $p_0[(x_1, x_2)] > p_1[(x_1, x_2)]$

COLUMBIA UNIVERSITY

Naïve Bayes Classifier (4)

- Gaussian Naïve Bayes - For quantitative features, the j-th feature in the k-th group is drawn from $N(\mu_{kj}, \sigma_{kj}^2)$
- Multinomial Naïve Bayes - For qualitative features, class-specific probabilities can be estimated from the contingency table.

	$Y = C_1$	$\cdots$	$Y = C_K$
$X = X_1$			
$\vdots$			
$X = X_J$			

Columbia University

Gaussian Naïve Bayes

- Gaussian Naïve Bayes - For quantitative features, the j-th feature in the k-th group is drawn from $N(\mu_{kj}, \sigma_{kj}^2)$
- The feature matrix is arranged as $X_{observation,feature}$ where x_{ij} is the j-th feature of the i-th observation

$$\hat{\pi}_k = \frac{n_k}{N}$$

$$\hat{\mu}_{kj} = \frac{1}{n_k} \sum_{i:y_i=k} x_{ij}$$

$$\hat{\sigma}_{kj}^2 = \frac{1}{n_k - 1} \sum_{i:y_i=k} \left(x_{ij} - \hat{\mu}_{kj}\right)^2$$

COLUMBIA UNIVERSITY

Multinomial Naïve Bayes

- Multinomial Naïve Bayes - For qualitative features, class-specific probabilities can be estimated from the contingency table.

$$\boldsymbol{y} = (y_1, \dots, y_K)^T,\ \sum_{k=1}^{K} y_k = n$$

$$f(\boldsymbol{y}|n) = \frac{n!}{y_1! \dots y_K!} \pi_1^{y_1} \dots \pi_K^{y_J}$$

COLUMBIA UNIVERSITY

That was

COLUMBIA UNIVERSITY