**PS5841**

# Data Science in Finance & Insurance

## Bayes Classifier

Yubo Wang

Autumn 2021

COLUMBIA UNIVERSITY

# Bayes Rule / Theorem

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

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

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

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

Columbia University

# 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

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

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

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

# 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)

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

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

Columbia University

# Gaussian Naïve Bayes (1)

- 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.

COLUMBIA UNIVERSITY

# Gaussian Naïve Bayes (2)

- For quantitative features
  - the feature matrix is arranged as

$$X_{observation,feature}$$

where $x_{ij}$ is the $j$-th feature of the $i$-th observation

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

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

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

COLUMBIA UNIVERSITY

# Multinomial Naïve Bayes

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

$$\boldsymbol{y} = \left(y_1, \ldots, y_J\right)^T, \ \sum_{j=1}^{J} y_j = n$$

$$f(\boldsymbol{y}|n) = \frac{n!}{y_1! \ldots y_J!} \pi_1^{y_1} \ldots \pi_J^{y_J}$$

COLUMBIA UNIVERSITY

# Example (1)
# 2 quantitative features & 2 classes

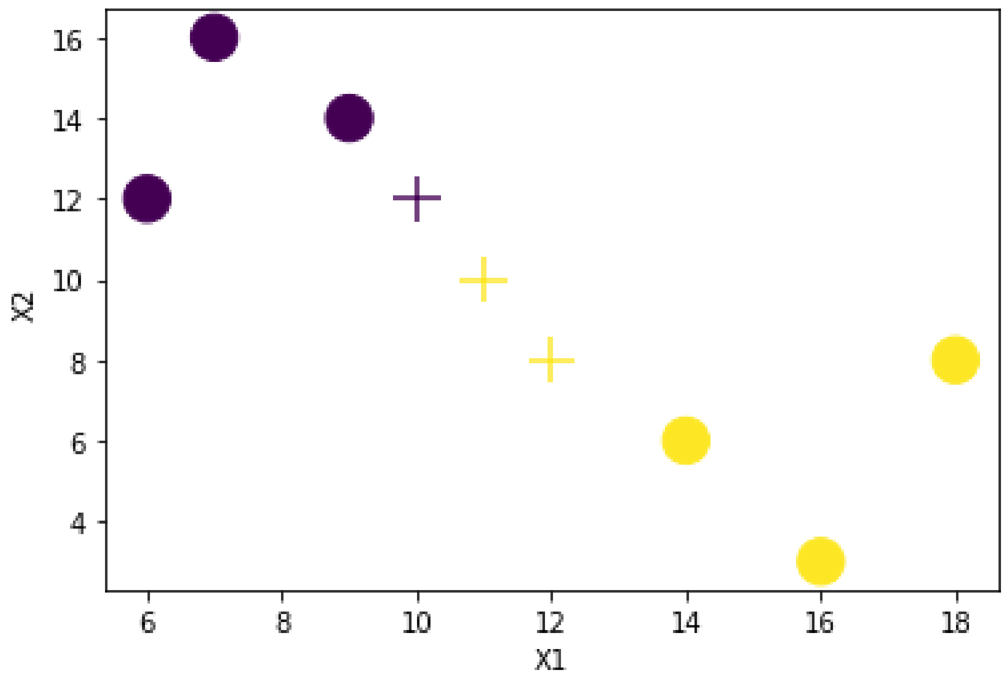
- Consider a particular observation or covariate pattern $\boldsymbol{x} = (x_1, x_2)$

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

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

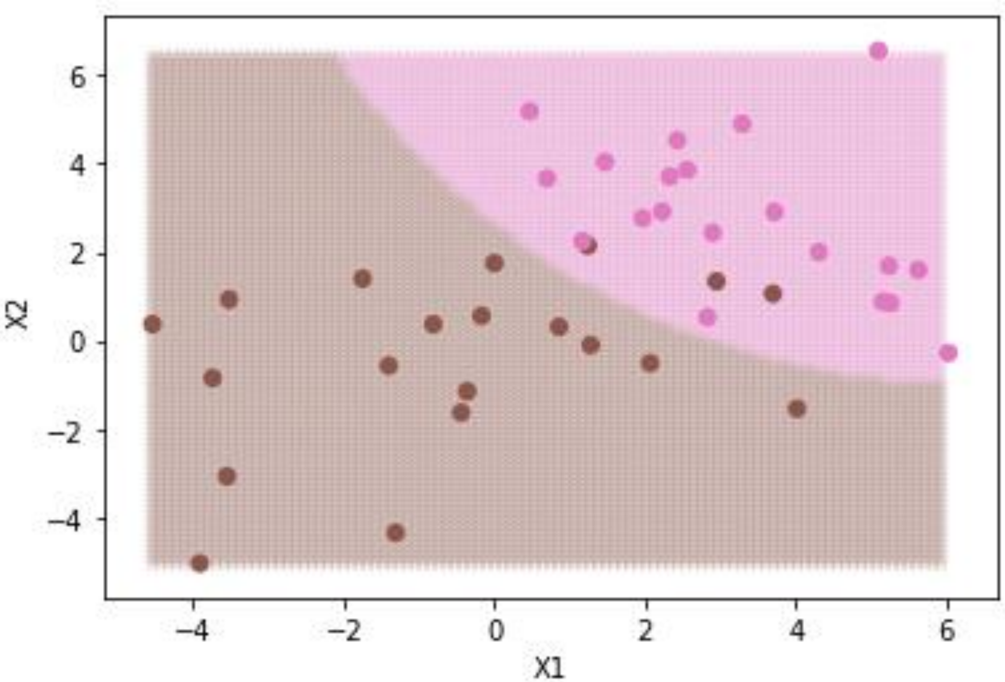
COLUMBIA UNIVERSITY

# Example (1)
# 2 quantitative features & 2 classes



# Decision Boundary

- 2-class response with 2D features



# Coding Stuff

- R

```
e1071: naiveBayes()
```

- Python

```
sklearn.naive_bayes.\

GaussianNB()
MultinomialNB()
BernoulliNB()
```

COLUMBIA UNIVERSITY

# That was