Naïve Bayes

Thursday, June 29, 2023 2:17 PM

Indep

Talks for more than 100 min? (TT >= 100)	Gender	Response
y	male	not bought
n	male	not bought
n	female	not bought
n	female	not bought
n	male	not bought
n	male	not bought
y	male	bought
y	female	bought
n	female	bought
y	female	bought

A & B indep

$\Rightarrow P(A \cap B \mid C) = P(A \mid C)\, P(B \mid C)$

$$P(B \mid TT \geq 100, m)$$

$$= \frac{P(TT \geq 100 \cap M \mid B)\, P(B)}{P(TT \geq 100 \cap M \mid B)\, P(B) + P(TT \geq 100 \cap M \mid NB)\, P(NB)}$$

$$= \frac{P(TT \geq 100 \mid B)\, P(M \mid B)\, P(B)}{P(TT \geq 100 \mid B)\, P(M \mid B)\, P(B) + P(TT \geq 100 \mid NB)\, P(M \mid NB)\, P(NB)}$$

Apriori Probabilities

$$= \frac{\frac{3}{4} \times \frac{1}{4} \times \frac{4}{10}}{\frac{3}{4} \times \frac{1}{4} \times \frac{4}{10} + \frac{1}{6} \times \frac{4}{6} \times \frac{6}{10}} = 0.529$$

Posterior Probability

$$P(B \mid TT \geq 100 \cap F) = \frac{P(TT \geq 100 \mid B)\, P(F \mid B)\, P(B)}{P(TT \geq 100 \mid B)\, P(F \mid B)\, P(B) + P(TT \geq 100 \mid NB)\, P(F \mid NB)\, P(NB)}$$

$$= \frac{\frac{3}{4} \times \frac{3}{4} \times \frac{4}{10}}{\frac{3}{4} \times \frac{3}{4} \times \frac{4}{10} + \frac{1}{6} \times \frac{2}{6} \times \frac{6}{10}} = 0.87$$

Data

Train → Calculating all the combinations of Apriori probs
• fit()

Test → y_test

Test → X_test

Calculating → X_test → Predictions
• predict()
• predict_proba()

y_test ↔ Predictions: Compared : acc

Discrete NB	Kernel NB

Categorical Features	Numerical Features
Apriori Probabilities are calculated based on counts	Apriori Probabilities are calculated based on the function of Normal Distribution
Bayes Formula	Bayes Formula

[Stratification | Why to Stratify? | stratify=y option]