

1 Naive Bayes for Text Categorization

Prior Probabilities :

N is the number of short documents = 8

$$\begin{aligned}\hat{P}(\text{vegetable}) &= \frac{N_{\text{veg}}}{N} \\ &= \frac{2}{8}\end{aligned}$$

$$\begin{aligned}\hat{P}(\text{fruit}) &= \frac{N_{\text{fruit}}}{N} \\ &= \frac{3}{8}\end{aligned}$$

$$\begin{aligned}\hat{P}(\text{flower}) &= \frac{N_{\text{flower}}}{N} \\ &= \frac{3}{8}\end{aligned}$$

Conditional Probabilities

$$\begin{aligned}\hat{P}(w|c) &= \frac{\text{Count}(w, c) + 1}{\text{Count}(c) + 1 * |V|} \\ &= \frac{\text{Count}(w, c) + 0.1}{\text{Count}(c) + 0.1 * |V|}\end{aligned}$$

Here $|V| = 14$

$$\hat{p}(\text{banana} | \text{vegetable}) = \frac{2+0.1}{8+1.4} = \frac{2.1}{9.4}$$

$$\hat{p}(\text{carrot} | \text{vegetable}) = \frac{1+0.1}{8+1.4} = \frac{1.1}{9.4}$$

$$\hat{p}(\text{cucumber} | \text{vegetable}) = \frac{1+0.1}{8+1.4} = \frac{1.1}{9.4}$$

$$\hat{p}(\text{pea} | \text{vegetable}) = \frac{2+0.1}{8+1.4} = \frac{2.1}{9.4}$$

$$\hat{p}(\text{potato} | \text{vegetable}) = \frac{1+0.1}{8+1.4} = \frac{1.1}{9.4}$$

$$\hat{p}(\text{baked} | \text{vegetable}) = \frac{1+0.1}{8+1.4} = \frac{1.1}{9.4}$$

$$\hat{p}(\text{lotus} | \text{vegetable}) = \frac{0+0.1}{8+1.4} = \frac{0.1}{9.4}$$

$$\hat{p}(\text{rose} | \text{vegetable}) = \frac{0+0.1}{8+1.4} = \frac{0.1}{9.4}$$

$$\hat{p}(\text{lily} | \text{vegetable}) = \frac{0+0.1}{8+1.4} = \frac{0.1}{9.4}$$

$$\hat{p}(\text{hibiscus} | \text{vegetable}) = \frac{0+0.1}{8+1.4} = \frac{0.1}{9.4}$$

$$\hat{p}(\text{grape} | \text{vegetable}) = \frac{0+0.1}{8+1.4} = \frac{0.1}{9.4}$$

$$\hat{p}(\text{school} | \text{vegetable}) = \frac{0+0.1}{8+1.4} = \frac{0.1}{9.4}$$

$$\hat{p}(\text{mango} | \text{vegetable}) = \frac{0+0.1}{8+1.4} = \frac{0.1}{9.4}$$

$$\hat{p}(\text{apple} | \text{vegetable}) = \frac{0+0.1}{8+1.4} = \frac{0.1}{9.4}$$

$$\hat{p}(\text{banana} | \text{fruit}) = \frac{1+0.1}{14+1.4} = \frac{1.1}{15.4}$$

$$\hat{p}(\text{carrot} | \text{fruit}) = \frac{0+0.1}{14+1.4} = \frac{0.1}{15.4}$$

$$\hat{p}(\text{cucumber} | \text{fruit}) = \frac{0+0.1}{14+1.4} = \frac{0.1}{15.4}$$

$$\hat{p}(\text{pea} | \text{fruit}) = \frac{0+0.1}{14+1.4} = \frac{0.1}{15.4}$$

$$\hat{p}(\text{potato} | \text{fruit}) = \frac{0+0.1}{14+1.4} = \frac{0.1}{15.4}$$

$$\hat{p}(\text{basket} | \text{fruit}) = \frac{0+0.1}{14+1.4} = \frac{0.1}{15.4}$$

$$\hat{p}(\text{lotus} | \text{fruit}) = \frac{1+0.1}{14+1.4} = \frac{1.1}{15.4}$$

$$\hat{p}(\text{rose} | \text{fruit}) = \frac{1+0.1}{14+1.4} = \frac{1.1}{15.4}$$

$$\hat{p}(\text{lily} | \text{fruit}) = \frac{1+0.1}{14+1.4} = \frac{1.1}{15.4}$$

$$\hat{p}(\text{hibiscus} | \text{fruit}) = \frac{2+0.1}{14+1.4} = \frac{2.1}{15.4}$$

$$\hat{p}(\text{grape} | \text{fruit}) = \frac{2+0.1}{14+1.4} = \frac{2.1}{15.4}$$

$$\hat{p}(\text{school} | \text{fruit}) = \frac{1+0.1}{14+1.4} = \frac{1.1}{15.4}$$

$$\hat{p}(\text{mango} | \text{fruit}) = \frac{2+0.1}{14+1.4} = \frac{2.1}{15.4}$$

$$\hat{p}(\text{apple} | \text{fruit}) = \frac{3+0.1}{14+1.4} = \frac{3.1}{15.4}$$

$$\hat{p}(\text{banana} | \text{flower}) = \frac{0+0.1}{13+1.4} = \frac{0.1}{14.4}$$

$$\hat{p}(\text{Carrot} | \text{flower}) = \frac{0+0.1}{13+1.4} = \frac{0.1}{14.4}$$

$$\hat{p}(\text{Cucumber} | \text{flower}) = \frac{1+0.1}{13+1.4} = \frac{1.1}{14.4}$$

$$\hat{p}(\text{pea} | \text{flower}) = \frac{2+0.1}{13+1.4} = \frac{2.1}{14.4}$$

$$\hat{p}(\text{potato} | \text{flower}) = \frac{0+0.1}{13+1.4} = \frac{0.1}{14.4}$$

$$\hat{p}(\text{basket} | \text{flower}) = \frac{0+0.1}{13+1.4} = \frac{0.1}{14.4}$$

$$\hat{p}(\text{lotus} | \text{flower}) = \frac{2+0.1}{13+1.4} = \frac{2.1}{14.4}$$

$$\hat{p}(\text{rose} | \text{flower}) = \frac{3+0.1}{13+1.4} = \frac{3.1}{14.4}$$

$$\hat{p}(\text{lily} | \text{flower}) = \frac{2+0.1}{13+1.4} = \frac{2.1}{14.4}$$

$$\hat{p}(\text{hibiscus} | \text{flower}) = \frac{3+0.1}{13+1.4} = \frac{3.1}{14.4}$$

$$\hat{p}(\text{grape} | \text{flower}) = \frac{0+0.1}{13+1.4} = \frac{0.1}{14.4}$$

$$\hat{p}(\text{school} | \text{flower}) = \frac{0+0.1}{13+1.4} = \frac{0.1}{14.4}$$

$$\hat{p}(\text{mango} | \text{flower}) = \frac{0+0.1}{13+1.4} = \frac{0.1}{14.4}$$

$$\hat{p}(\text{apple} | \text{flower}) = \frac{0+0.1}{13+1.4} = \frac{0.1}{14.4}$$

Posterior Probabilities :

$$P(c|d) = \frac{P(d|c) P(c)}{P(d)}$$

Naïve Bayes Classifier

$$C_{MAP} = \underset{c \in C}{\operatorname{argmax}} P(d|c) P(c)$$

Choosing a class

$$P(\text{vegetable} | D1) \propto \frac{2}{8} \times \frac{0.1}{9.4} \times \frac{0.1}{9.4} \times \frac{0.1}{9.4} \times \frac{1.1}{9.4}$$
$$\approx 3.52225893e-8$$

$$P(\text{fruit} | D1) \propto \frac{3}{8} \times \frac{1.1}{15.4} \times \frac{1.1}{15.4} \times \frac{3.1}{15.4} \times \frac{0.1}{15.4}$$
$$\approx 0.0000250089$$

$$P(\text{flower} | D1) \propto \frac{3}{8} \times \frac{3.1}{14.4} \times \frac{2.1}{14.4} \times \frac{0.1}{14.4} \times \frac{0.1}{14.4}$$
$$\approx 5.67756726e-7$$

Most likely class for D1 is fruit.

$$P(\text{vegetable} | D2) \propto \frac{2}{8} \times \frac{2.1}{9.4} \times \frac{1.1}{9.4} \times \frac{0.1}{9.4} \times \frac{0.1}{9.4}$$

$$\approx 7.3967e^{-7}$$

$$P(\text{fruit} | D2) \propto \frac{3}{8} \times \frac{0.1}{15.4} \times \frac{0.1}{15.4} \times \frac{1.1}{15.4} \times \frac{2.1}{15.4}$$

$$\approx 1.5401e^{-7}$$

$$P(\text{flower} | D2) \propto \frac{3}{8} \times \frac{2.1}{14.4} \times \frac{0.1}{14.4} \times \frac{2.1}{14.4} \times \frac{0.1}{14.4}$$

$$\approx 3.8460e^{-7}$$

Most likely class for D2 is vegetable