The Hong Kong Polytechnic University
Department of Applied Mathematics

AMA4680 Statistical Machine Learning

Naive Bayes Classifier

1. (Naive Bayes classifier with Laplacian smoothing) Given the following 15 training samples

Name	Give Birth	Lay Eggs	Can Fly	Have Legs	Class
monkey	yes	no	no	yes	mammal
whale	yes	no	no	no	mammal
bat	yes	no	yes	yes	mammal
cat	yes	no	no	yes	mammal
dolphin	yes	no	no	no	mammal
python	no	yes	no	no	non-mammal
salmon	no	yes	no	no	non-mammal
frog	no	yes	no	yes	non-mammal
lizard	no	yes	no	yes	non-mammal
pigeon	no	yes	yes	yes	non-mammal
leopard shark	yes	no	no	no	non-mammal
turtle	no	yes	no	yes	non-mammal
penguin	no	yes	no	yes	non-mammal
owl	no	yes	yes	yes	non-mammal
eagle	no	yes	yes	yes	non-mammal

predict the label of the following sample $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$

Name	Give Birth	Lay Eggs	Can Fly	Have Legs	Class
human	yes	no	no	yes	?

Solution. The goal is to compare

$$\mathbb{P}(\mathsf{mammal})\Pi_{i=1}^{4}\mathbb{P}(x_i|\mathsf{mammal}) \quad \text{and} \quad \mathbb{P}(\mathsf{nonmammal})\Pi_{i=1}^{4}\mathbb{P}(x_i|\mathsf{nonmammal})$$

Step 1: compute prior probabilities

$$\mathbb{P}(C = \mathsf{mammal}) = \frac{5+1}{15+2} = \frac{6}{17}$$

and

$$\mathbb{P}(C = \mathsf{nonmammal}) = \frac{10+1}{15+2} = \frac{11}{17}$$

Step 2: compute likelihoods.

- (mammal)

$$\mathbb{P}(\mathsf{Give\ birth} = \mathsf{yes}|C = \mathsf{mammal}) = \frac{N_{\mathsf{mammal,Give\ birth=yes}}+\alpha}{N_{\mathsf{mammal}}+\alpha\times(\#\mathsf{Give\ birth})} = \frac{5+1}{5+2} = \frac{6}{7},$$

$$\mathbb{P}(\mathsf{Lay\ eggs} = \mathsf{no}|C = \mathsf{mammal}) = \frac{N_{\mathsf{mammal,Lay\ eggs=no}}+\alpha}{N_{\mathsf{mammal}}+\alpha\times(\#\mathsf{Lay\ eggs})} = \frac{5+1}{5+2} = \frac{6}{7},$$

$$\mathbb{P}(\mathsf{Can\ fly} = \mathsf{no}|C = \mathsf{mammal}) = \frac{N_{\mathsf{mammal,Can\ fly=no}}+\alpha}{N_{\mathsf{mammal}}+\alpha\times(\#\mathsf{Can\ fly})} = \frac{4+1}{5+2} = \frac{5}{7},$$

$$\mathbb{P}(\mathsf{Have\ legs} = \mathsf{yes}|C = \mathsf{mammal}) = \frac{N_{\mathsf{mammal,Have\ legs=yes}}+\alpha}{N_{\mathsf{mammal}}+\alpha\times(\#\mathsf{Have\ legs})} = \frac{3+1}{5+2} = \frac{4}{7}.$$

- (non-mammal)

$$\mathbb{P}(\textsf{Give birth} = \textsf{yes}|C = \textsf{nonmammal}) = \frac{N_{\textsf{nonmammal,Give birth=yes}}+\alpha}{N_{\textsf{nonmammal}}+\alpha\times(\#\textsf{Give birth})} = \frac{1+1}{10+2} = \frac{1}{6},$$

$$\mathbb{P}(\textsf{Lay eggs} = \textsf{no}|C = \textsf{nonmammal}) = \frac{N_{\textsf{nonmammal,Lay eggs=no}}+\alpha}{N_{\textsf{nonmammal}}+\alpha\times(\#\textsf{Lay eggs})} = \frac{1+1}{10+2} = \frac{1}{6},$$

$$\mathbb{P}(\textsf{Can fly} = \textsf{no}|C = \textsf{nonmammal}) = \frac{N_{\textsf{nonmammal,Can fly=no}}+\alpha}{N_{\textsf{nonmammal}}+\alpha\times(\#\textsf{Can fly})} = \frac{7+1}{10+2} = \frac{2}{3},$$

$$\mathbb{P}(\textsf{Have legs} = \textsf{yes}|C = \textsf{nonmammal}) = \frac{N_{\textsf{nonmammal,Have legs=yes}}+\alpha}{N_{\textsf{nonmammal}}+\alpha\times(\#\textsf{Have legs})} = \frac{7+1}{10+2} = \frac{2}{3}.$$

Step 3: comparison.

$$\begin{aligned} & \frac{6}{17}\times\frac{6}{7}\times\frac{6}{7}\times\frac{5}{7}\times\frac{4}{7}\approx 0.1058 \\ > & \frac{11}{17}\times\frac{1}{6}\times\frac{1}{6}\times\frac{2}{3}\times\frac{2}{3}\approx 0.0052 \end{aligned}$$

Step 4: conclusion. the sample $\boldsymbol{x}$ belongs to mammal.

2. (Naive Bayes classifier without Laplacian smoothing) Consider the following dataset about weather conditions and whether people played outside:

Outlook	**Temperature**	**Humidity**	**Wind**	**Play**
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rain	Mild	High	False	Yes
Rain	Cool	Normal	False	Yes
Rain	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rain	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rain	Mild	High	True	No

Predict the label of the sample $\boldsymbol{x} = (x_1, x_2, x_3, x_4) = (\mathsf{Sunny}, \mathsf{Hot}, \mathsf{High}, \mathsf{False})$.

Solution.

Step 1: Calculate Prior Probabilities

Calculate the prior probabilities for each class (*Play* = *Yes*, *Play* = *No*).

- Total instances = 14
- Instances where Play = Yes: 9
- Instances where Play = No: 5

Prior Probabilities:

$$P(Play = Yes) = \frac{9}{14} \approx 0.643$$

$$P(Play = No) = \frac{5}{14} \approx 0.357$$

Step 2: Calculate Likelihoods

Calculate the likelihoods for each feature given the class label.

For $Play = Yes$

- **Outlook**:
  - Sunny: 2
  - Overcast: 4
  - Rain: 3
- **Temperature**:
  - Hot: 2
  - Mild: 4
  - Cool: 3
- **Humidity**:
  - High: 3

- Normal: 6
- **Windy**:
  - True: 3
  - False: 6

Likelihood Probabilities:

$$P(Outlook = Sunny|Play = Yes) = \frac{2}{9}$$

$$P(Temperature = Hot|Play = Yes) = \frac{2}{9}$$

$$P(Humidity = High|Play = Yes) = \frac{3}{9}$$

$$P(Wind = True|Play = Yes) = \frac{3}{9}$$

For $Play = No$

- **Outlook**:
  - Sunny: 3
  - Overcast: 0
  - Rain: 2
- **Temperature**:
  - Hot: 2
  - Mild: 1
  - Cool: 2
- **Humidity**:
  - High: 4
  - Normal: 1
- **Windy**:
  - True: 2
  - False: 3

Likelihood Probabilities:

$$P(Outlook = Sunny|Play = No) = \frac{3}{5}$$

$$P(Temperature = Hot|Play = No) = \frac{2}{5}$$

$$P(Humidity = High|Play = No) = \frac{4}{5}$$

$$P(Wind = True|Play = No) = \frac{2}{5}$$

Step 3: Classify the new sample

Calculate Posterior Probability for Each Class

For $Play = Yes$

$$
\begin{aligned}
& P(Yes) \cdot P(Outlook = Sun|Yes) \cdot P(Temp = Hot|Yes) \cdot P(Hum = High|Yes) \cdot P(Wind = False|Yes) \\
\approx\ & 0.643 \cdot \frac{2}{9} \cdot \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{6}{9} \\
\approx\ & 0.643 \cdot 0.02469 \approx 0.01585
\end{aligned}
$$

For $Play = No$

$$
\begin{aligned}
& P(No) \cdot P(Outlook = Sun|No) \cdot P(Temp = Hot|No) \cdot P(Hum = High|No) \cdot P(Wind = False|No) \\
\approx\ & 0.357 \cdot \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} \\
\approx\ & 0.357 \cdot 0.096 \approx 0.03429
\end{aligned}
$$

We classify the new sample as **"No"** (not playing outside) since $0.01585 < 0.03429$.