NAIVE BAYES CLASSIFIER

Supervised learning algorithm based on the application of Bayes' theorem with the "naive" assumption of independence between every pair of features.

1 Bayes' Theorem

$$\begin{aligned}\mathbf{P}(A \cap B) &= \mathbf{P}(A|B) \times \mathbf{P}(B) \\ &= \mathbf{P}(B|A) \times \mathbf{P}(A)\end{aligned}$$

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(B|A) \times \mathbf{P}(A)}{\mathbf{P}(B)}$$

2 Derivation

$$\mathbf{P}(\mathbf{Y} = y_k \mid x_1,\ x_2,\ ...\ x_p) = \frac{\mathbf{P}(x_1,\ x_2,\ ...\ x_p|\mathbf{Y} = y_k) \times \mathbf{P}(\mathbf{Y} = y_k)}{\mathbf{P}(x_1,\ x_2,\ ...\ x_p)}$$

Choose y_k which maximizes $\mathbf{P}(\mathbf{Y} = y_k|x_1,\ x_2,\ ...\ x_p)$.

$$\mathbf{Y} = \underset{y_k}{\operatorname{argmax}} \left\{ \ \tfrac{\mathbf{P}(x_1,\ x_2,\ ...\ x_p|\mathbf{Y}=y_k)\times\mathbf{P}(\mathbf{Y}=\mathbf{y_k})}{\mathbf{P}(x_1,\ x_2,\ ...\ x_p)} \right.$$

Since $\mathbf{P}(x_1,\ x_2,\ ...\ x_p)$ is constant for all y_k, the above can be written as follows:

$$\mathbf{Y} = \underset{y_k}{\operatorname{argmax}} \left\{ \ \mathbf{P}(x_1,\ x_2,\ ...\ x_p|\mathbf{Y} = y_k) \times \mathbf{P}(\mathbf{Y} = \mathbf{y_k}) \right.$$

The "naive" assumption that all $\mathbf{X_1},\ \mathbf{X_2},\ ...\ \mathbf{X_p}$ are independent allows for the following simplification.

$$\mathbf{Y} = \underset{y_k}{\operatorname{argmax}} \{ \prod_i \mathbf{P}(x_i|\mathbf{Y} = y_k) \times \mathbf{P}(\mathbf{Y} = \mathbf{y_k})$$

All the terms on the right are easy to calculate.

3 Example

Table 1: Training data

Sr.	**Outlook**	**Temperature**	**Humidity**	**Windy**	**PlayGolf**
1	Rainy	Hot	High	False	No
2	Rainy	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Sunny	Mild	High	False	Yes
5	Sunny	Cool	Normal	False	Yes
6	Sunny	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Rainy	Mild	High	False	No
9	Rainy	Cool	Normal	False	Yes
10	Sunny	Mild	Normal	False	Yes
11	Rainy	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Sunny	Mild	High	True	No

The following tables are calculated.

Outlook	**Yes**	**No**
Rainy	2	3
Overcast	4	0
Sunny	3	2

Temperature	**Yes**	**No**
Hot	2	2
Mild	4	2
Cool	3	1

Humidity	**Yes**	**No**
High	3	4
Normal	6	1

Windy	**Yes**	**No**
False	6	2
True	3	3

If $X = (Sunny, Hot, Normal, False)$, then

$$\textbf{PlayGolf} = \operatorname*{argmax}_{y_k} \begin{cases} \mathbf{P}(\textbf{Outlook} = Sunny|\textbf{PlayGolf} = Yes) \\ \times\mathbf{P}(\textbf{Temprature} = Hot|\textbf{PlayGolf} = Yes) \\ \times\mathbf{P}(\textbf{Humidity} = Normal|\textbf{PlayGolf} = Yes) \\ \times\mathbf{P}(\textbf{Windy} = False|\textbf{PlayGolf} = Yes) \\ \times\mathbf{P}(\textbf{PlayGolf} = Yes) \\ \\ \mathbf{P}(\textbf{Outlook} = Sunny|\textbf{PlayGolf} = No) \\ \times\mathbf{P}(\textbf{Temprature} = Hot|\textbf{PlayGolf} = No) \\ \times\mathbf{P}(\textbf{Humidity} = Normal|\textbf{PlayGolf} = No) \\ \times\mathbf{P}(\textbf{Windy} = False|\textbf{PlayGolf} = No) \\ \times\mathbf{P}(\textbf{PlayGolf} = No) \end{cases}$$

$$\textbf{PlayGolf} = \operatorname*{argmax}_{y_k} \begin{cases} 3/9 \\ \times 2/9 \\ \times 6/9 \\ \times 6/9 \\ \times 9/14 \\ \\ 2/5 \\ \times 2/5 \\ \times 1/5 \\ \times 2/5 \\ \times 5/14 \end{cases}$$

$$\textbf{PlayGolf} = \operatorname*{argmax}_{y_k} \begin{cases} 0.0212 \\ 0.0046 \end{cases}$$

Therefore, $\textbf{PlayGolf} = Yes$ because $0.0212 > 0.0046$.

4 Extensions

- For a continuous input feature, assumption regarding the distribution needs to be made. Examples: **Gaussian**, **Multinomial** and **Bernoulli**.
- Smoothing may be required to prevent the multiplication from being zero when one probability term is zero.

5 Comments

- Due to independence assumption, naive Bayes' classifiers often perform good even with less training data.

- Main applications include **spam filtering** and **document classification**.
- Extremely fast in both training and prediction.
- Often fail to produce a good estimate of the correct class probabilities but make the correct classification if the correct class is more probable than any other class.