Tutorial 7

Tutorial on Naive Bayes

Ben Halstead, Luke Chang

The University of Auckland

May 2021

Topics

Example 1 - Multinomial Naive Bayes Classifier

Example

You come to Fiji for a holiday. 10 days later, you realise the weather forecast here isn't very accurate. Based on the information you gettered so far and today's weather report, you want to know "Will it rain this afternoon?"

Day	**Outlook** (O)	**Temperature** (T)	**Humidity** (H)	**Wind** (W)	**Rain** (R)
1	Sunny	Hot	High	Weak	True
2	Sunny	Hot	High	Strong	False
3	Overcast	Hot	High	Weak	True
4	Rain	Mild	High	Weak	True
5	Rain	Cool	Normal	Weak	True
6	Rain	Cool	Normal	Strong	False
7	Overcast	Cool	Normal	Strong	False
8	Overcast	Mild	High	Strong	True
9	Sunny	Cool	Normal	Weak	False
10	Rain	Mild	Normal	Weak	False
11	Sunny	Mild	Normal	Strong	?

Example 1 - Formulate the Problem

- Attribute: *Outlook* (O), *Temperature* (T), *Humidity* (H), *Wind* (W).
- Output: *Rain* (R) - Binary classification problem
- How can we formulate this task?

Example 1 - Formulate the Problem

- Attribute: *Outlook* (O), *Temperature* (T), *Humidity* (H), *Wind* (W).
- Output: *Rain* (R) - Binary classification problem
- How can we formulate this task?
  - The probability of a given output $r \in \{\text{True}, \text{False}\}$ is:

$$P(R = r | O, T, H, W)$$

Example 1 - Formulate the Problem

- Attribute: *Outlook* (O), *Temperature* (T), *Humidity* (H), *Wind* (W).
- Output: *Rain* (R) - Binary classification problem
- How can we formulate this task?
  - The probability of a given output $r \in \{\text{True}, \text{False}\}$ is:

$$P(R = r | O, T, H, W)$$

  - We want to predict the label with the highest probability.

$$R = \underset{r \in \{\text{T}, \text{F}\}}{\text{argmax}} \, P(R = r | O, T, H, W)$$

Example 1 - Formulate the Problem

- Bayes Theorem:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

- Terminology – Prior: $P(Y)$, Likelihood: $P(X|Y)$, Posterior: $P(Y|X)$, Marginal Probability: $P(X)$.
- How to rewrite this expression using Bayes Theorem?

Example 1 - Formulate the Problem

- Bayes Theorem:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

- Terminology – Prior: $P(Y)$, Likelihood: $P(X|Y)$, Posterior: $P(Y|X)$, Marginal Probability: $P(X)$.
- How to rewrite this expression using Bayes Theorem?

$$R = \operatorname*{argmax}_{r \in \{\mathsf{T},\mathsf{F}\}} P(R = r|O,T,H,W)$$

$$R = \operatorname*{argmax}_{r \in \{\mathsf{T},\mathsf{F}\}} \frac{P(O,T,H,W|R = r)P(R = r)}{P(O,T,H,W)}$$

- How can we simplify this problem?

Example 1 - Formulate the Problem

- Bayes Theorem:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

- Terminology – Prior: $P(Y)$, Likelihood: $P(X|Y)$, Posterior: $P(Y|X)$, Marginal Probability: $P(X)$.
- How to rewrite this expression using Bayes Theorem?

$$R = \operatorname*{argmax}_{r\in\{\mathsf{T},\mathsf{F}\}} P(R=r|O,T,H,W)$$

$$R = \operatorname*{argmax}_{r\in\{\mathsf{T},\mathsf{F}\}} \frac{P(O,T,H,W|R=r)P(R=r)}{P(O,T,H,W)}$$

- How can we simplify this problem?
$P(Y|X)$ is proportional to: $P(Y|X) \propto P(X|Y)P(Y)$

$$R = \operatorname*{argmax}_{r\in\{\mathsf{T},\mathsf{F}\}} P(O,T,H,W|R=r)P(R=r)$$

- The marginal Probability $P(O,T,H,W)$ is omitted. If we want to know the probability, we can normalise all possible outcomes.

Example 1 - Calculate the Prior $P(R)$

- 10 observations
- 5 days was raining.
- 5 days wasn't.

Example 1 - Calculate the Prior $P(R)$

- 10 observations
- 5 days was raining.
- 5 days wasn't.
- $P(R = \text{True}) = \frac{5}{10} = 0.5$
- $P(R = \text{False}) = \frac{5}{10} = 0.5$

Example 1 - Calculate the Prior $P(R)$

- 10 observations
- 5 days was raining.
- 5 days wasn't.
- $P(R = \text{True}) = \frac{5}{10} = 0.5$
- $P(R = \text{False}) = \frac{5}{10} = 0.5$

Example

Gambler's fallacy: Toss a coin, I see 20 head showed up in a row. The next time it must land on tail, since the probability of a coin landing 21 times is too low, 0.5^{21}.

Example 1 - Calculate the Prior $P(R)$

- 10 observations
- 5 days was raining.
- 5 days wasn't.
- $P(R = \text{True}) = \frac{5}{10} = 0.5$
- $P(R = \text{False}) = \frac{5}{10} = 0.5$

Example

Gambler's fallacy: Toss a coin, I see 20 head showed up in a row. The next time it must land on tail, since the probability of a coin landing 21 times is too low, 0.5^{21}.

Problem:

- Frequentist Statistics: Each tail is independent.
- $P(Y) \neq P(Y|X)$
- Bayes Theorem: If the same coin is used, the more evidence you collect, the more likely the coin is loaded.

Example 1 - Calculate the Likelihood

- The first part of the formula, $P(O,T,H,W|R=r)$, is the *likelihood*.
- It describes how *likely* an event occurs, given the outcome.
- How do we calculate the *likelihood*?

Example 1 - Calculate the Likelihood

- The first part of the formula, $P(O,T,H,W|R=r)$, is the *likelihood*.
- It describes how *likely* an event occurs, given the outcome.
- How do we calculate the *likelihood*?
- We could try to calculate $P(O,T,H,W|R=r)$ directly. What is the problem with this approach?
  - We need to compute all possible combinations.
  - We have: $3 \times 3 \times 2 \times 2 = 36$ different combinations of O,T,H,W *per label*.
  - We only have 10 observations – not enough to calculate probabilities for all combinations.
- Naive Bayes makes the assumption that all attributes are independent: allowing us to calculate the likelihood as:

$$P(O,T,H,W|R=r) = P(O|R=r)p(T|R=r)p(H|R=r)p(W|R=r)$$

Example 1 - Calculate the Likelihood

Day	Outlook (O)	Temperature (T)	Humidity (H)	Wind (W)	Rain (R)
1	Sunny	Hot	High	Weak	True
3	Overcast	Hot	High	Weak	True
4	Rain	Mild	High	Weak	True
5	Rain	Cool	Normal	Weak	True
8	Overcast	Mild	High	Strong	True

$$\sum P(X = x_i | R = r) = 1$$

- $P(O = \text{Sunny} | R = \text{T}) = \frac{1}{5}$
- $P(O = \text{Overcast} | R = \text{T}) = \frac{2}{5}$
- $P(O = \text{Rain} | R = \text{T}) = \frac{2}{5}$

- $P(H = \text{High} | R = \text{T}) = \frac{4}{5}$
- $P(H = \text{Normal} | R = \text{T}) = \frac{1}{5}$

- $P(T = \text{Hot} | R = \text{T}) = \frac{2}{5}$
- $P(T = \text{Mild} | R = \text{T}) = \frac{2}{5}$
- $P(T = \text{Cool} | R = \text{T}) = \frac{1}{5}$

- $P(T = \text{Strong} | R = \text{T}) = \frac{1}{5}$
- $P(T = \text{Weak} | R = \text{T}) = \frac{4}{5}$

Example 1 - Calculate the Likelihood

Day	Outlook (O)	Temperature (T)	Humidity (H)	Wind (W)	Rain (R)
2	Sunny	Hot	High	Strong	False
6	Rain	Cool	Normal	Strong	False
7	Overcast	Cool	Normal	Strong	False
9	Sunny	Cool	Normal	Weak	False
10	Rain	Mild	Normal	Weak	False

- $P(O = \text{Sunny}|R = \text{F}) = \frac{2}{5}$
- $P(O = \text{Overcast}|R = \text{F}) = \frac{1}{5}$
- $P(O = \text{Rain}|R = \text{F}) = \frac{2}{5}$
- $P(H = \text{High}|R = \text{F}) = \frac{1}{5}$
- $P(H = \text{Normal}|R = \text{F}) = \frac{4}{5}$
- $P(T = \text{Hot}|R = \text{F}) = \frac{1}{5}$
- $P(T = \text{Mild}|R = \text{F}) = \frac{1}{5}$
- $P(T = \text{Cool}|R = \text{F}) = \frac{3}{5}$
- $P(T = \text{Strong}|R = \text{F}) = \frac{3}{5}$
- $P(T = \text{Weak}|R = \text{F}) = \frac{2}{5}$

Example 1 - Calculate the Posterior

Day	**Outlook** (O)	**Temperature** (T)	**Humidity** (H)	**Wind** (W)	**Rain** (R)
11	Sunny	Mild	Normal	Strong	?

Now we have:

$$R = \underset{r \in \{\mathsf{T},\mathsf{F}\}}{\operatorname{argmax}} P(O|R=r)P(T|R=r)P(H|R=r)P(W|R=r)P(R=r)$$

$$\begin{aligned} P(R=\mathsf{T}|O,T,H,W) &\propto P(O=\mathsf{S}|R=\mathsf{T})P(T=\mathsf{M}|R=\mathsf{T})P(H=\mathsf{N}|R=\mathsf{T})P(W=\mathsf{S}|R=\mathsf{T})P(R=\mathsf{T}) \\ &\propto \frac{1}{5}\cdot\frac{2}{5}\cdot\frac{1}{5}\cdot\frac{1}{5} = 0.0032 \end{aligned}$$

$$\begin{aligned} P(R=\mathsf{F}|O,T,H,W) &\propto P(O=\mathsf{S}|R=\mathsf{F})P(T=\mathsf{M}|R=\mathsf{F})P(H=\mathsf{N}|R=\mathsf{F})P(W=\mathsf{S}|R=\mathsf{F})P(R=\mathsf{F}) \\ &\propto \frac{2}{5}\cdot\frac{1}{5}\cdot\frac{4}{5}\cdot\frac{3}{5} = 0.0384 \end{aligned}$$

Example 1 - How Likely?

Day	Outlook (O)	Temperature (T)	Humidity (H)	Wind (W)	Rain (R)
11	Sunny	Mild	Normal	Strong	?

$$P(R = \mathsf{T}|O,T,H,W) = \frac{0.0032}{P(O,T,H,W)}$$

$$P(R = \mathsf{F}|O,T,H,W) = \frac{0.0384}{P(O,T,H,W)}$$

$$P(R = \mathsf{T}|O,T,H,W) = \frac{0.0032}{0.0032+0.0384} = 0.08$$

$$P(R = \mathsf{F}|O,T,H,W) = \frac{0.0384}{0.0032+0.0384} = 0.92$$

Given the information, there is 92% to not rain.