RZR

Naive Bayes

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

⇒ From the table above find whether a player will play tennis ($\to P$) or not ($\to \neg P$): when outlook is sunny (S), temperature is cool (C), humidity is high (h) and wind is strong (St).

$\Rightarrow$ Basically you need to find;

$P(P \mid S \wedge C \wedge H \wedge St)$

$P(\neg P \mid S \wedge C \wedge H \wedge St)$

Find these two values and compare.

Ans

Considering bayes theorem:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

$$\Rightarrow P(P \mid S \wedge C \wedge H \wedge St) = \frac{P(S \wedge C \wedge H \wedge St \mid P)\, P(P)}{P(S \wedge C \wedge H \wedge St)}$$

$$\Rightarrow P(\neg P \mid S \wedge C\, H \wedge St) = \frac{P(S \wedge C \wedge H \wedge St \mid \neg P)\, P(\neg P)}{P(S \wedge C \wedge H \wedge St)}$$

* While comparing the denominator will have the same impact on both; so we omit that and compare the values of numerator only.

→ let's calculate for probability of playing tennis:

$\Rightarrow P(P \mid S \cap C \cap H \cap St)$

$= P(S \cap C \cap H \cap St \mid P)\ P(P)$

$= P(S \mid P) \times P(C \mid P) \times P(H \mid P) \times P(St \mid P) \times P(P)$

$= \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14} = 0.0053$

⇒ Now calculating for not playing tennis:

$P(\neg P \mid S \cap C \cap H \cap St)$

$= P(S \cap C \cap H \cap St \mid \neg P)\ P(\neg P)$

$= P(S \mid \neg P) \times P(C \mid \neg P) \times P(H \mid \neg P) \times P(St \mid \neg P) \times P(\neg P)$

$= \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14} = 0.0206.$

So, the player will not likely to play tennis.

⇒ Now let's say you are asked to solve the same problem by for outlook is overcast; and all the other conditions are same as before.

⇒ to solve this at some point you will need to find the probability $P(\text{overcast} \mid \neg P)$

↳ No.

* But if you look to the table; there is no overcast data for playing tennis is no.

↳ Hence the probability will be 0 overall. But that's a problem. So, we need to consider "Laplace Correction".

⇒ For that;

we add 1 with all the data for No.

$\Rightarrow$ So;

Sunny $= 3+1 \rightarrow P(S|\neg P) = \frac{4}{8}$

Rainy $= 2+1 \rightarrow P(R|\neg P) = \frac{3}{8}$

Overcast $= 0+1 \rightarrow \underline{P(Oc|\neg P)} = \frac{1}{8}$

$\hookrightarrow$ this is not 0 anymore.

$\Rightarrow$ So the probability will not be 0.

⇒ What if we have to deal with numeric values?

⇒ let's say the temperature column contains numeric values instead.

For yes:

25.2, 19.3, 18.5, 21.7, 20.1, 24.3, 22.8, 23.1, 19.8

For no:

27.3, 30.1, 17.4, 29.5, 15.1

⇒ What if you need to find → $P(P \mid S \cap 32 \mid H \cap St)$?
temp

$$\Rightarrow P(S \mid P)\, P(32 \mid P)\, P(H \mid P)\, P(Ht \mid P)\, P(P)$$

$$\Rightarrow \frac{2}{9} \times P(32 \mid P) \times \frac{3}{5} \times \frac{3}{5} \times \frac{9}{14}$$

$$P(32 \mid P) \rightarrow \frac{1}{\sqrt{2\pi}\,\sigma_y} \times e^{\frac{-(x-\mu_y)^2}{2 \times (\sigma_y)^2}}$$

Here; $x \rightarrow 32$ (temp value that we are working on)

$\sigma_y \rightarrow$ Sigma (Standard Deviation)

$\mu_y \rightarrow$ Mean (Average of the data points for yes)

$$\Rightarrow \sigma_y = \sqrt{\frac{\sum_{i=1}^{NDP} (\text{Datapoint} - \text{Mean value})^2}{\text{Number of data points}}}$$

$\Rightarrow \mu_y$ ↓

$$\frac{\text{Sum}(25.2,\ 19.3,\ 18.5,\ 21.7,\ 20.1,\ 24.3,\ 22.8,\ 23.1,\ 19.8)}{9}$$

$= 21.64.$

$$\text{So, } \sigma_y = \sqrt{\frac{(25.2 - 21.64)^2 + (19.3 - 21.64)^2 + \cdots\cdots}{9}}$$

$= 2.35$

$$P(32|P) = \frac{1}{\sqrt{2\pi} \times \sigma_y} \times e^{-\frac{(x-\mu_y)^2}{2\times(\sigma_y)^2}}$$

$$= \frac{1}{\sqrt{2\times 3.1416} \times 2.35} \times e^{-\frac{(32 - 21.64)^2}{2\times(2.35)^2}}$$

$= \boxed{}$ → Complete the rest. R2R