100 Days of ML

Naive Bayes Classifier

Conditional Probability

Events A & B

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, given $P(B) \neq 0$

→ Probability of A provided B has already happened.

Eg - Two dices thrown D_1 & D_2

$\{(1,1), (1,2), \dots, (1,6), (2,1), (2,2), \dots, (2,6), \dots, (6,6)\}$

$\Rightarrow 36$

D_2 \ D_1	1	2	3	4	5	6
1	2	3	4	5	6 ✓	7
2	3	4	5	6	7 ✓	8
3	4	5	6	7	8 ✓	9
4	5	6	7	8	9 ✓	10
5	6	7	8	9	10 ✓	11
6	7	8	9	10	11	12

i) Probability of $D_1 = 5$

$\Rightarrow P(A : D_1 = 5) = \dfrac{1}{36}$

ii) Probality of $D_1 + D_2 \leq 10$

$\Rightarrow P(A : D_1 + D_2 \leq 10) = \dfrac{33}{36} = \dfrac{11}{12}$

iii) Probability of $D_1 = 5$ given $D_1 + D_2 \leq 10$

$\Rightarrow P(\underset{A}{D_1 = 5} \mid \underbrace{D_1 + D_2 \leq 10}_{B}) = \dfrac{P(D_1 = 5 \cap D_1 + D_2 \leq 10)}{P(D_1 + D_2 \leq 10)}$

$= \dfrac{5/36}{33/36} = \dfrac{5}{33}$

Independent Events

$$P(A \cap B) = P(A) \times P(B)$$

$$\Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \times \cancel{P(B)}}{\cancel{P(B)}} = P(A)$$

Mutually Exclusive Events

$$P(A \cap B) = 0$$

$$\Rightarrow P(A|B) = 0$$

Bayes Theorem

$$\underbrace{P(A|B)}_{\text{Posterior}} = \frac{\overbrace{P(B|A)}^{\text{Likelihood}} \times \underbrace{P(A)}_{\text{Prior}}}{\underbrace{P(B)}_{\text{Evidence}}}, \text{ given } P(B) \neq 0$$

$$\Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{--- (1)}$$

$$A \cap B = B \cap A$$

$$\Rightarrow P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)}$$

$$\Rightarrow P(A \cap B) = P(B|A) \times P(A) \quad \text{--- (2)}$$

From (1) & (2),

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

Q. Factory

M_1	M_2	M_3	
20	30	50	— Units
5%	3%	1%	— Defective prob.

Page No.
Date

Find probability of marker being picked from M_3 given that it is defective.

$$\Rightarrow P(M_1) = \frac{1}{5} \qquad P(M_2) = \frac{3}{10} \qquad P(M_3) = \frac{1}{2}$$

$$P(D|M_1) = \frac{1}{20} \qquad P(D|M_2) = \frac{3}{100} \qquad P(D|M_3) = \frac{1}{100}$$

ATQ: $P(M_3|D)$

$$\Rightarrow P(M_3|D) = \frac{P(D|M_3) \times P(M_3)}{P(D)}$$

$$P(D) = P(D \cap M_1) + P(D \cap M_2) + P(D \cap M_3)$$
$$= P(D|M_1) \times P(M_1) + P(D|M_2) \times P(M_2) + P(D|M_3) \times P(M_3)$$
$$= \frac{1}{20} \times \frac{1}{5} + \frac{3}{100} \times \frac{3}{10} + \frac{1}{100} \times \frac{1}{2}$$

Naive Bayes Intuition

Toss	Venue	Outlook	Result
Won	Mumbai	Overcast	Won
Lost	Chennai	Sunny	Won
Won	Kolkata	Sunny	Won
Won	Chennai	Sunny	Won
Lost	Mumbai	Sunny	Lost
Won	Chennai	Overcast	Lost
Won	Kolkata	Overcast	Lost
Won	Mumbai	Sunny	Won

CSK

Predict if CSK win or lose for {Lost, Mumbai, Sunny}

$P(W \mid Lost \cap Mumbai \cap Sunny)$
$P(L \mid Lost \cap Mumbai \cap Sunny)$

$$\Rightarrow \frac{P(Lost \cap Mumbai \cap Sunny \mid W)\, P(W)}{P(Lost \cap Mumbai \cap Sunny)}$$

According to Naive Bayes:

$$\frac{P(Lost \mid W) \times P(Mumbai \mid W) \times P(Sunny \mid W) \times P(W)}{P(Lost \cap Mumbai \cap Sunny)}$$ ← Don't require as D^n is same for both Win or Lose prediction

$$= (1/5)(2/5)(4/5)(5/8)$$

$$= \frac{1}{25} = 0.04$$

IIIly, $$\frac{P(Lost \cap Mumbai \cap Sunny \mid L)\, P(L)}{P(Lost \cap Mumbai \cap Sunny)}$$

$$\Rightarrow P(Lost \mid L) \times P(Mumbai \mid L) \times P(Sunny \mid L) \times P(L)$$

$$= 1/3 \times 1/3 \times 1/3 \times 3/8$$

$$= \frac{1}{72} = 0.013$$

∴ CSK will win

Numerical Data in Naive Bayes

Height	Weight	Gender
172	150	M
180	170	M
165	140	M
190	200	M
139	100	F
145	120	F
160	140	F
172	150	F

Predict Gender, given H=185 & W=170

Page No.
Date

<u>Assumption</u>: Height & Weight are a Gaussian distributed random ~~value~~ variable

⇒ Assuming that Height & Weight are normally distributed

$P(M | H=185, W=170)$

$= P(H=185|M) \times P(W=170|M) \times \underbrace{P(M)}_{1/2}$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

(Normal distribution eqn ; μ = mean, σ = standard deviation)

$P(F | H=185, W=170)$

$= P(H=185| F) \times P(W=170|F) \times \underbrace{P(F)}_{1/2}$