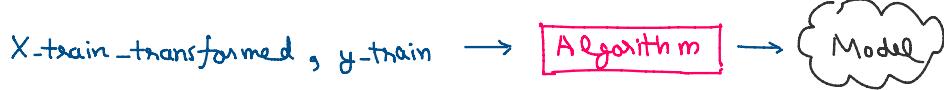


Building a Machine Learning Model

1. Training



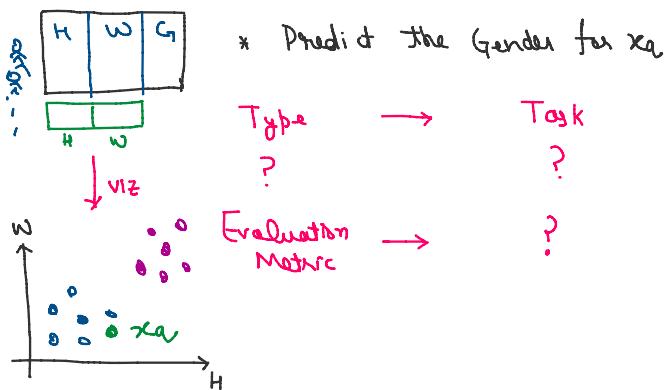
2. Predicting using unseen data



3. Evaluation



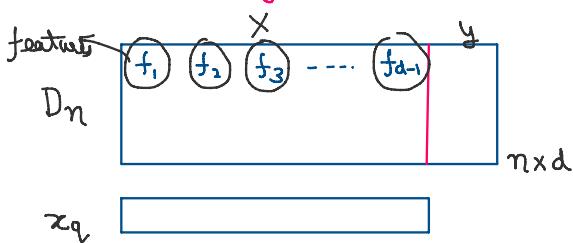
- Given Data containing Height, Weights & Gender of some individuals.



- Target Variable → Gender (y)
 $\text{Gender} \in \{M, F\}$
- Task → Classification
- Evaluation Metrics → Accuracy
(How?) Probabilistic Approach
- Algorithms → Naïve Bayes Algorithm

Naïve Bayes Algorithm ↴ (AKA Maximum A Posteriori)

Type - Supervised Learning
Task - Classification.



Given $\rightarrow D_n = \{(x_i, y_i)\}_{i=1}^n \mid x_i \in \mathbb{R}^{d-1}, y_i \in \{c_1, c_2, \dots, c_k\}$

$$x_q = f_1 \cap f_2 \cap \dots \cap f_{d-1}$$

Task → Find y_q
How? → Given x_q , compute the prob of x_q belonging to each of the classes.
Whichever is maximum, assign that class to x_q .

$$\text{i.e. } P(c_1 | x_q) \cdot P(c_2 | x_q) \cdot \dots \cdot P(c_k | x_q)$$

$$y_q = \arg \max_{c_1, c_2, \dots, c_k} \left\{ \frac{P(c_i | x_q)}{\text{Posteriori Prob.}} \right\}$$

From Bayes Theorem ↴

$$\frac{\text{Likelyhood}}{\text{Prior}}$$

From Bayes Theorem \rightarrow

$$P(C_i | x_q) = \frac{\underbrace{P(x_q | C_i)}_{\text{Posterior}} * \underbrace{P(C_i)}_{\text{Prior}}}{P(x_q) \underset{\text{Marginal}}{]}}$$

Key Concepts \rightarrow

\rightarrow Probability (Study of uncertainty)

Random Experiment

- It is a process for which outcome cannot be predicted with certainty.

Example \rightarrow

R.E \rightarrow Tossing a coin

Sample Space

- It is a set of all possible outcomes of a 'RANDOM EXPERIMENT'.

Example \rightarrow

S.S = {H, T}

Event

- It is a subset of a 'SAMPLE SPACE'.

Example \rightarrow

Event - Getting a Head
of H}

\rightarrow Probability \rightarrow

$$P(\text{Event}) = \frac{\text{No. of favourable Outcomes}}{\text{Total no. of Outcomes}} = \frac{|\text{Event}|}{|\text{Sample Space}|}$$

\rightarrow Example-1 :

R.E \rightarrow Tossing 2 coins

S.S \rightarrow {HH, HT, TH, TT}

Event A,

A : Getting atleast 1 Head

A \rightarrow {HH, HT, TH}

$$P(A) = 3/4$$

\rightarrow Example-2 :

R.E \rightarrow Rolling a dice

S.S \rightarrow {1, 2, 3, 4, 5, 6}

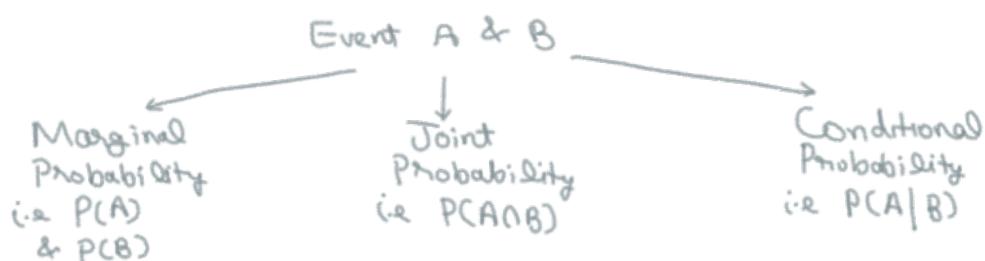
Event B,

B : Getting a Prime No.

B \rightarrow {2, 3, 5}

$$P(B) = 3/6$$

\rightarrow In case of two or more events \rightarrow



* Next we will see how to compute these probabilities

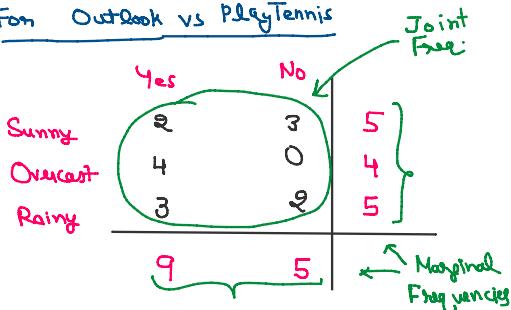
- * Next we will see how to compute these probabilities with the help of 'FREQUENCY TABLE'.

Consider this binary classification data set:

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

Frequency Table

For Outlook vs PlayTennis



Marginal Prob →

$$P(\text{PlayTennis} = \text{Yes}) = \frac{9}{14}$$

$$P(\text{PlayTennis} = \text{No}) = \frac{5}{14}$$

$$P(\text{Outlook} = \text{Sunny}) = \frac{9}{14}$$

Joint Prob →

$$P(\text{Outlook} = \text{Sunny} \cap \text{PlayTennis} = \text{Yes}) = \frac{2}{14}$$

$$P(\text{Outlook} = \text{Overcast} \cap \text{PlayTennis} = \text{No}) = \frac{0}{14}$$

$$P(\text{Outlook} = \text{Rainy} \cap \text{PlayTennis} = \text{No}) = \frac{2}{14}$$

Conditional Prob →

$$P(\text{Outlook} = \text{Sunny} \mid \text{PlayTennis} = \text{Yes}) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{2/14}{9/14} = \frac{2}{9}$$

Bayes Theorem →

$$\text{Acc}^n \rightarrow \text{Conditional Prob} \rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} \quad -\textcircled{1}$$

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

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

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

$$P(A \cap B) = P(B|A)P(A) \quad -\textcircled{2}$$

Substituting $\textcircled{2}$ in eqⁿ $\textcircled{1}$:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \begin{array}{l} \text{Likelihood} \\ \downarrow \\ \text{Posterior} \end{array} \quad \begin{array}{l} \text{Prior} \\ \leftarrow \\ \text{Marginal} \end{array}$$

Independent Events \rightarrow

If given events A & B are known to be independent:

$$P(A|B) = P(A)$$

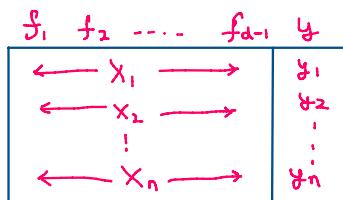
$$P(B|A) = P(B)$$

Similarly:

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

Data \rightarrow

$$D_n = \left\{ (x_i, y_i) \mid x_i \in \mathbb{R}^{d-1}, y_i \in \{0, 1\} \right\}$$



$$x_i = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{d-1} \end{bmatrix}$$

Naive Bayes Algorithm

Task \rightarrow Classification

How?

Given x_q , compute $P(c_1|x_q), P(c_2|x_q), \dots, P(c_k|x_q)$.

Whichever probability is maximum, prediction would be the corresponding class.

$$\hat{y}_q = \arg \max_{i \in \{1, 2, \dots, k\}} \{P(c_i|x_q)\} \quad \text{ie maximum a posteriori}$$

$$\rightarrow P(c_1|x_q) = \frac{P(x_q|c_1) P(c_1)}{P(x_q)} \Rightarrow P(c_1|x_q) \propto P(x_q|c_1) P(c_1)$$

$$\rightarrow P(c_2|x_q) = \frac{P(x_q|c_2) P(c_2)}{P(x_q)} \Rightarrow P(c_2|x_q) \propto P(x_q|c_2) P(c_2)$$

⋮

⋮

$$\rightarrow P(c_k|x_q) = \frac{P(x_q|c_k) P(c_k)}{P(x_q)} \Rightarrow P(c_k|x_q) \propto P(x_q|c_k) P(c_k)$$

Naive Bayes Derivation \rightarrow

Given $x_q = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{d-1} \end{bmatrix}$ i.e. x_q is $f_1 \cap f_2 \cap f_3 \dots \cap f_{d-1}$

Given $X_q = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{d-1} \end{bmatrix}$ i.e. X_q is $f_1 \cap f_2 \cap f_3 \dots \cap f_{d-1}$

Using Bayes Theorem:

$$P(C_k | x_q) = \frac{P(x_q | C_k) * P(C_k)}{P(x_q)}$$

$$\text{Compute} \rightarrow P(c_k | x_q) \propto P(x_q | c_k) P(c_k)$$

$$\hat{y}_q = \arg \max_{k \in \{1, 2, \dots, K\}} \left\{ P(C_k | x_q) \right\}$$

$$\hat{y}_q = \arg \max_{k \in \{1, 2, \dots, K\}} \left\{ p(x_q | C_k) p(C_k) \right\}$$

↳ Let's zoom inside this term and understand how to compute it step by step.

According to Conditional Probability 7

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

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

$P(X_q | C_k) P(C_k)$ can be written as $P(X_q \cap C_k)$

We have already seen that x_0 is $f_1 \cap f_2 \cap \dots \cap f_{d-1}$

$$\rightarrow P(x_q | c_k) P(c_k)$$

$$\rightarrow P(x_a \cap c_k)$$

$$\rightarrow P(f_1 \cap f_2 \cap f_3 \dots f_{d-1} \cap c_k)$$

$$P(f_1 \cap f_2 \cap f_3 \cap \dots \cap f_{d-1} \cap C_k)$$

Using $P(A \cap B) = P(A|B)P(B)$

$$\rightarrow P(f_1 \mid f_2 \cap f_3 \dots \cap f_{d-1} \cap C_k) *$$

$$P(f_1 \cap f_2 \cap f_3 \dots \cap f_{d-1} \cap C_k)$$

Expanding this expressing using
Chain Rule of conditional Probability

$$\rightarrow P(f_1 | f_2 \cap f_3 \dots \cap f_{d-1} \cap C_k) * \\ P(f_2 | f_3 \cap f_4 \dots \cap f_{d-1} \cap C_k) * \\ \vdots \\ P(f_{d-1} | C_k) * P(C_k)$$

\therefore Above expression is extremely hard to compute.

To simplify the computation lets take a
NAIVE assumption that:

Features are conditionally independent
of each other.

i.e if $f_1, f_2, f_3, \dots, f_{d-1}$ are independent
of each other,

$$P(f_1 | f_2 \cap f_3 \dots \cap f_{d-1} \cap C_k) = P(f_1 | C_k)$$

$$\rightarrow P(f_1 | C_k) * P(f_2 | C_k) * \dots * P(f_{d-1} | C_k) * P(C_k)$$

$$\rightarrow P(C_k) * \prod_{i=1}^{d-1} P(f_i | C_k)$$

$$\hat{y}_q = \arg \max_{J \in \{1, 2, \dots, k\}} \left\{ \begin{array}{c} P(x_q | C_J) P(C_J) \\ \downarrow \\ P(C_J) * \prod_{i=1}^{d-1} P(f_i | C_J) \end{array} \right\}$$

$$\text{i.e } \hat{y}_q = \arg \max_{J \in \{1, 2, \dots, k\}} \left\{ P(C_J) * \prod_{i=1}^{d-1} P(f_i | C_J) \right\}$$