Course: MCA

Semester: IV

Paper Code/Name: DSE4T2 (Introduction to Machine Learning)

Topic: Naïve Bayes Algorithm

Faculty Name: Dr. Tapan Kant

Email: tapan.kant@gmail.com

Bayes Theorem

- P (h) = prior probability of hypothesis h
- P (D) = prior probability of training data D
- P (h|D) = probability of h given D
- P (D|h) = probability of D given h

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Roles of Bayesian Method

- Provides practical learning algorithms:
  - Naive Bayes learning
  - Bayesian belief network learning
  - Combine prior knowledge (prior probabilities) with observed data
  - Requires prior probabilities
- Provides useful conceptual framework
  - Provides \gold standard" for evaluating other learning algorithms
  - Additional insight into Occam's razor

What is Naïve Bayes Classifier?

- It is a statistical classification technique based on Bayes Theorem
- A simplest supervised learning algorithms
- It is fast, accurate and reliable algorithm
- It has high accuracy and speed on large datasets.
- Application – Spam filtering, text classification, sentiment analysis and recommender system.

- It assumes that the effect of a particular feature in a class is independent of other features
- Loan Applicant – income, previous loan history, transaction history, age, location, etc.
- This assumption is called class conditional independence

Naive Bayes Algorithm

Naive_Bayes_Learn($examples$)

For each target value v_j

$\hat{P}(v_j) \leftarrow$ estimate $P(v_j)$

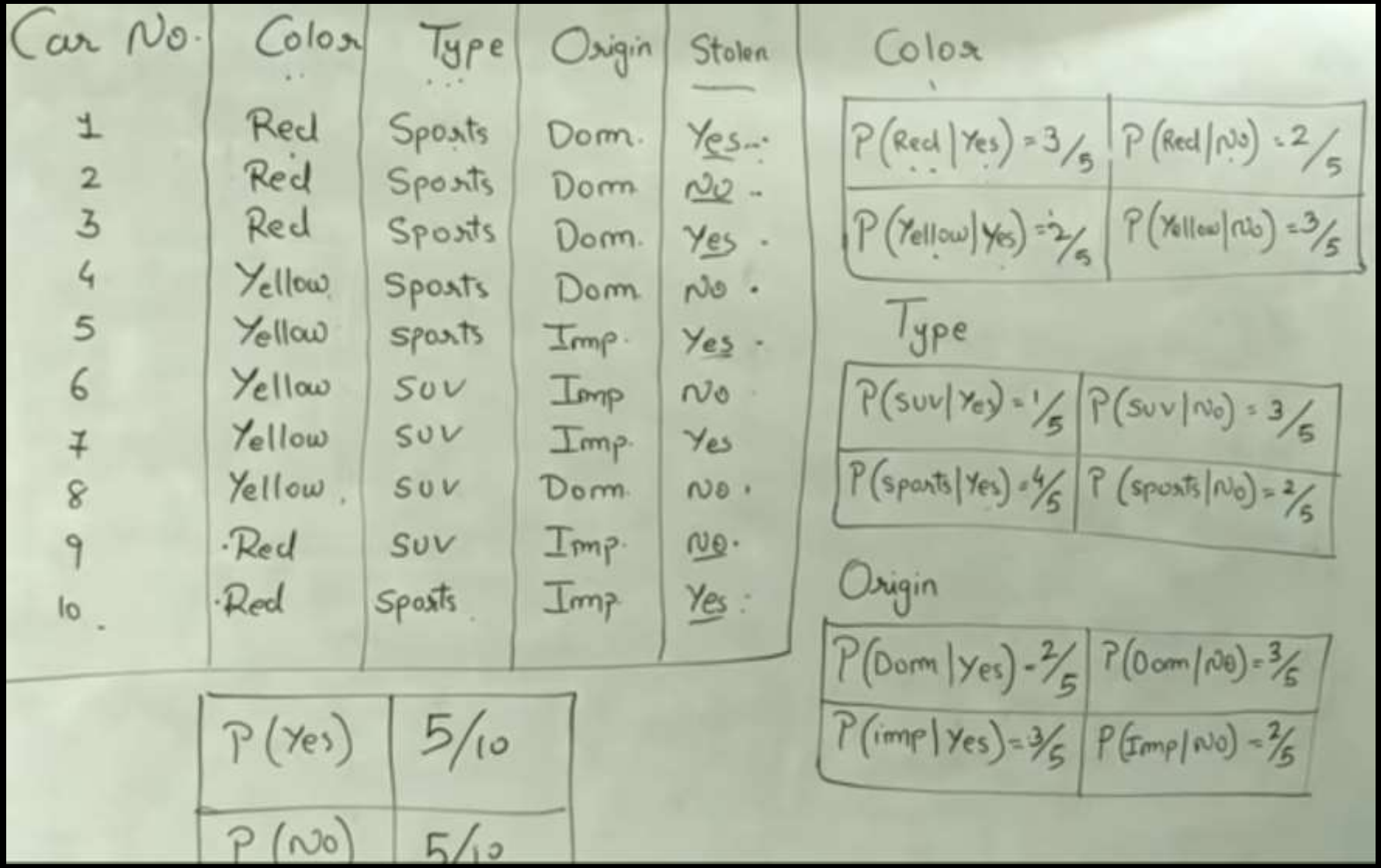
For each attribute value a_i of each attribute a

$\hat{P}(a_i|v_j) \leftarrow$ estimate $P(a_i|v_j)$

Classify_New_Instance(x)

$$v_{NB} = \underset{v_j \in V}{\operatorname{argmax}} \hat{P}(v_j) \prod_{a_i \in x} \hat{P}(a_i|v_j)$$

Example



Car No.	Color	Type	Origin	Stolen
1	Red	Sports	Dom.	Yes
2	Red	Sports	Dom	No
3	Red	Sports	Dom.	Yes
4	Yellow	Sports	Dom	No
5	Yellow	sports	Imp	Yes
6	Yellow	SUV	Imp	No
7	Yellow	SUV	Imp	Yes
8	Yellow	SUV	Dom	No
9	Red	SUV	Imp	No
10	Red	Sports	Imp	Yes

P(Yes)	5/10
P(No)	5/10

Color

P(Red | Yes) = 3/5	P(Red | No) = 2/5
P(Yellow | Yes) = 2/5	P(Yellow | No) = 3/5

Type

P(SUV | Yes) = 1/5	P(SUV | No) = 3/5
P(sports | Yes) = 4/5	P(sports | No) = 2/5

Origin

P(Dom | Yes) = 2/5	P(Dom | No) = 3/5
P(Imp | Yes) = 3/5	P(Imp | No) = 2/5

Sample $X = \langle \text{Red}, \text{SUV}, \text{Domestic} \rangle$

$$P(X \mid \text{Yes}) \cdot P(\text{Yes}) = P(\text{Red} \mid \text{Yes}) \cdot P(\text{SUV} \mid \text{Yes}) \cdot P(\text{Dom} \mid \text{Yes}) \cdot P(\text{Yes})$$

$$P(X \mid \text{No}) \cdot P(\text{No}) = P(\text{Red} \mid \text{No}) \cdot P(\text{SUV} \mid \text{No}) \; P(\text{Dom} \mid \text{No}) \; P(\text{No})$$

Sample $X = \langle \text{Red}, \text{SUV}, \text{Domestic} \rangle$

$$P(X \mid \text{Yes}) \cdot P(\text{Yes}) = P(\text{Red} \mid \text{Yes}) \cdot P(\text{SUV} \mid \text{Yes}) \cdot P(\text{Dom} \mid \text{Yes}) \cdot P(\text{Yes})$$

$$= 3/5 \cdot 1/5 \cdot 2/5$$

$$P(X \mid \text{No}) \cdot P(\text{No}) = P(\text{Red} \mid \text{No}) \cdot P(\text{SUV} \mid \text{No}) \cdot P(\text{Dom} \mid \text{No}) \cdot P(\text{No})$$

$$2/5 \cdot 3/5 \cdot 3/5$$

Sample $X = \langle \text{Red}, \text{SUV}, \text{Domestic} \rangle$

$$P(X \mid \text{Yes}) \cdot P(\text{Yes}) = P(\text{Red} \mid \text{Yes}) \cdot P(\text{SUV} \mid \text{Yes}) \cdot P(\text{Dom} \mid \text{Yes}) \cdot P(\text{Yes})$$

$$= 3/5 \cdot 1/5 \cdot 2/5$$

$$= 0.024$$

$$P(X \mid \text{No}) \cdot P(\text{No}) = P(\text{Red} \mid \text{No}) \cdot P(\text{SUV} \mid \text{No}) \cdot P(\text{Dom} \mid \text{No}) \cdot P(\text{No})$$

$$2/5 \cdot 3/5 \cdot 3/5$$

$$= 0.072$$

Thank you