Bayesian Classifier

Dr. Singara Singh Kasana

Associate Professor

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology

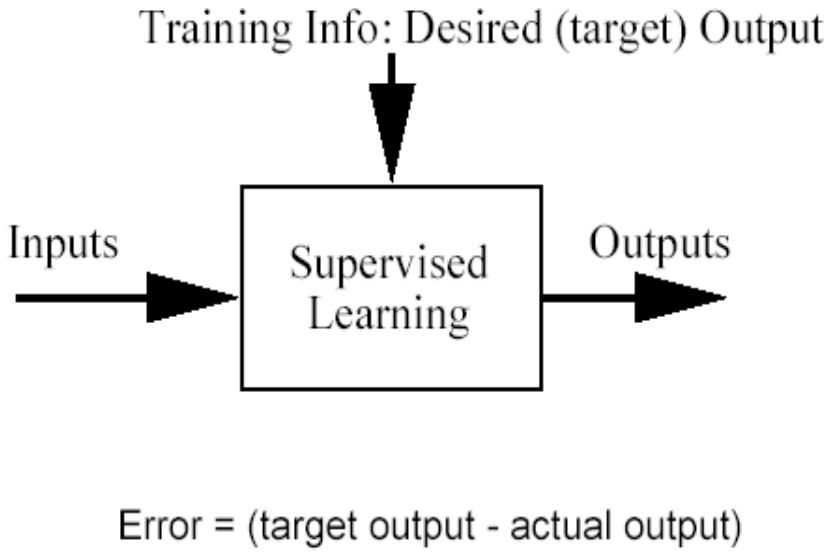
Patiala, Punjab

Bayesian Classifier

- Probabilities Classifier
- Supervised Machine Learning Model
- Fast and easy to implement
- Naïve in nature

Classification problem

- Training data: examples of the form (d,h(d))
  - where d are the data objects to classify (inputs)
  - and h(d) gives the class info for d, $h(d) \in \{1,\ldots K\}$
- Goal: given d_{new}, provide $h(d_{new})$



Assumption

The fundamental Naive Bayes assumption is that each feature makes an:

- independent
- equal

contribution to the output

Dataset

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

Assumption

With relation to our dataset, this concept can be understood as:

- We assume that no pair of features are dependent. For example, the temperature being 'Hot' has nothing to do with the humidity or the outlook being 'Rainy' has no effect on the winds. Hence, the features are assumed to be **independent**.
- Secondly, each feature is given the same weight(or importance). For example, knowing only temperature and humidity alone can't predict the outcome accurately. None of the attributes is irrelevant and assumed to be contributing **equally** to the outcome.

Example : Naïve Bayes

Predict playing tennis in the day with the condition <sunny, cool, high, strong> (P(v| o=sunny, t= cool, h=high w=strong)) using the following training data:

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

Naive Bayesian Classifier

Given a training set, we can compute the probabilities

Outlook	P	N
sunny	2/9	3/5
overcast	4/9	0
rain	3/9	2/5
Tempreature		
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humidity	P	N
high	3/9	4/5
normal	6/9	1/5
Windy		
true	3/9	3/5
false	6/9	2/5

Play-tennis example: estimating $P(x_i|C)$

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

P(y) = 9/14
P(n) = 5/14

outlook	
P(sunny|p) = 2/9	**P(sunny|n) = 3/5**
P(overcast|p) = 4/9	**P(overcast|n) = 0**
P(rain|p) = 3/9	**P(rain|n) = 2/5**
temperature	
P(hot|p) = 2/9	**P(hot|n) = 2/5**
P(mild|p) = 4/9	**P(mild|n) = 2/5**
P(cool|p) = 3/9	**P(cool|n) = 1/5**
humidity	
P(high|p) = 3/9	**P(high|n) = 4/5**
P(normal|p) = 6/9	**P(normal|n) = 1/5**
windy	
P(true|p) = 3/9	**P(true|n) = 3/5**
P(false|p) = 6/9	**P(false|n) = 2/5**

Predict the class of the day with the condition <sunny, cool, high, strong>

Let X= <sunny, cool, high, strong>

We need to calculate P(Yes|X) and P(No|X)

For these, we use Bayes theorem

P(Yes|X) = P(Yes) * P(X|Yes) /P(X)

P(No|X) = P(No) * P(X|No) /P(X)

P(Yes| outlook=sunny, temp= cool, humidity=high wind=strong) by using the given training data:

p(yes) = 9/14

p(outlook = sunny|yes) = 2/9

p(temp = cool|yes) = 3/9

p(humidity = high|yes) = 3/9

p(wind = strong|yes) = 3/9

9/14 * 2/9 * 3/9 * 3/9* 3/9 = 0.0053

Calculate P(No| outlook=sunny, temp= cool, humidity=high wind=strong) using the given training data:

p(No) = 5/14

p(outlook = sunny|no) = 3/5

p(temp = cool|no) = 1/5

p(humidity = high|no) = 4/5

p(wind = strong|no) = 3/5

5/14 * 3/5 * 1/5 * 4/5* 3/5 = 0.0206

P(x) = P(outlook=sunny) * P(Temperature = cool) * P(humidity = high)*P(Wind = strong)

= 5/14* 4/14*7/14*6/14 = 0.02186

P(Play= Y| x) = 0.0053/0.02186 = 0.2424

P(Play =N|x) = 0.0206/0.02186 = 0.9421

Thanks