Nguyen Vu Anh Ngoc
Machine Learning 1
October 9, 2023

Homework Week 7: Naive Bayes

Question 1. Coding: Implement Gaussian Naive Bayes

Question 2. Naive Bayes Example

No	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

1. **Estimate conditional probabilities of each attributes color, legs, height, smelly for the species classes:** $\{M, H\}$ **using the data given in the table.**

Attribute	P(Attribute|M)	P(Attribute|H)
Color=White	0.5	0.75
Color=Green	0.5	0.25
Legs=3	0.75	0.0
Legs=2	0.25	1.0
Height=Short	0.75	0.5
Height=Tall	0.25	0.5
Smelly=Yes	0.75	0.25
Smelly=No	0.25	0.75

2. **Using these probabilities estimate the probability values for the new instance - (Color=Green, legs=2, Height=Tall, and Smelly=No)**

Probability values for the species classes M and H using the Naive Bayes formula:

$$\begin{aligned} P(M|\text{Attributes}) \propto & P(M)\times \\ & P(\text{Color=Green}|M) \times P(\text{Legs=2}|M)\times \\ & P(\text{Height=Tall}|M) \times P(\text{Smelly=No}|M) \\ P(H|\text{Attributes}) \propto & P(H)\times \\ & P(\text{Color=Green}|H) \times P(\text{Legs=2}|H)\times \\ & P(\text{Height=Tall}|H) \times P(\text{Smelly=No}|H) \end{aligned}$$

To normalize the probabilities so that they sum to 1, we can use:

$$P(M|\text{Attributes}) = \frac{P(M|\text{Attributes})}{P(M|\text{Attributes}) + P(H|\text{Attributes})}$$

$$P(H|\text{Attributes}) = \frac{P(H|\text{Attributes})}{P(M|\text{Attributes}) + P(H|\text{Attributes})}$$

Given the new instance with attributes Color=Green, Legs=2, Height=Tall, Smelly=No:

$$P(M|\text{New Instance}) = 0.0769 < P(H|\text{New Instance}) = 0.9231$$

It's much more likely that this new instance belongs to species **H**.

Question 3. Naive Bayes for Continuous Data

Based on the following data determine the gender of a person having height 6 ft., weight 130 lbs, and foot size 8 inch.

Person	Height (ft)	Weight (lbs)	Foot size (Inches)
Male	6.00	180	12
Male	5.92	190	11
Male	5.58	170	12
Male	5.92	165	10
Female	5.00	100	6
Female	5.50	150	8
Female	5.42	130	7
Female	5.75	150	9

The prior probabilities for each gender are:

- $P(\text{Male}) = 0.5$ - $P(\text{Female}) = 0.5$

The estimated means and variances for each attribute given each gender are:

Attribute	**Male**	**Female**
Mean Height (ft)	5.855	5.4175
Variance in Height	0.035033	0.097225
Mean Weight (lbs)	176.25	132.5
Variance in Weight	122.916667	558.333333
Mean Foot Size (inches)	11.25	7.5
Variance in Foot Size	0.916667	1.666667

With these values, we can now compute the conditional probabilities for a person with the attributes Height = 6 ft., Weight = 130 lbs, Foot Size = 8 inches using the Gaussian distribution formula:

$$P(x|\text{Gender}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

The estimated probabilities for the given instance with attributes Height = 6 ft., Weight = 130 lbs, Foot Size = 8 inches are:

$$P(\text{Male}) \approx 1.33 \times 10^{-10} << P(\text{Female}) \approx 99.9999999867$$

It's likely that the person with the given attributes is **Female**.