

Travis Browning
University of Wolverhampton
Machine Learning, L7N008
Dr. Burcu Can
Assignment 1

For task 1, I began by calculating by hand which music genre Jane was most likely to enjoy. After introducing smoothing, it appeared most likely that Jane would enjoy pop music based on Euclidean distance. I chose the kNN classifier provided by scikit-learn for comparison. The classifier reliably predicted that Jane would be interested in Rock music. I found this surprising. According to my calculations summarized in table 1, Jane's 4th nearest neighbor was her nearest neighbor interested in rock music.

Table 1

Row Labels	Sum of Euc(Jane)
Scarlett	1.414
Jazz	
Hazel	2
Pop	
Ken	2.236
Pop	
Mary	2.236
Rock	
Lily	2.236
Jazz	
Tony	2.236
Rock	
Evelyn	2.236
Pop	

James	2.449
Rock	
Audrey	3.162
Pop	
Mason	3.317
Jazz	

For task 2, it appeared after calculating conditional probability by hand, that rock and pop were Jane's most likely preferred genres. I chose scikit-learn again for a Naive Bayes classifier. 11 unique features were found, while 30 features were used. Similarly to the performance of kNN, NB found the genre that I found least likely, in this case jazz, as the predicted preference.

Table 2

	Name	Features	Class
Training	James	20, M, Germany	Rock
	Tony	22, M, Germany	Rock
	Mason	30, M, UK	Jazz
	Mary	20, F, Germany	Rock
	Scarlett	26, F, UK	Jazz
	Ken	18, M, Spain	Pop
	Lily	30, F, UK	Jazz
	Audrey	16, F, UK	Pop
	Hazel	18, F, Spain	Pop
	Evelyn	18, M, Spain	Pop
Test	Jane	22, F, Spain	???

TASK 1

Euclidean distance(Jane):

James:

$$\begin{aligned}&\sqrt{(2-3)^2 + (0-2)^2 + (0-1)^2} \\&\sqrt{(-1)^2 + (-2)^2 + (-1)^2} \\&\sqrt{1 + 4 + 1} \\&\sqrt{6} = 2.449\end{aligned}$$

Tony:

$$\begin{aligned}&\sqrt{(3-3)^2 + (0-2)^2 + (0-1)^2} \\&\sqrt{(0)^2 + (-2)^2 + (-1)^2} \\&\sqrt{0 + 4 + 1} \\&\sqrt{5} = 2.236\end{aligned}$$

Mason:

$$\begin{aligned}&\sqrt{(5-3)^2 + (1-2)^2 + (0-1)^2} \\&\sqrt{(3)^2 + (-1)^2 + (-1)^2} \\&\sqrt{9 + 1 + 1} \\&\sqrt{11} = 3.317\end{aligned}$$

Mary:

$$\begin{aligned}&\sqrt{(2-3)^2 + (0-2)^2 + (1-1)^2} \\&\sqrt{(-1)^2 + (-2)^2 + (0)^2} \\&\sqrt{1 + 4 + 0} \\&\sqrt{5} = 2.236\end{aligned}$$

Scarlett:

$$\begin{aligned}&\sqrt{(4-3)^2 + (1-2)^2 + (1-1)^2} \\&\sqrt{(1)^2 + (-1)^2 + (0)^2} \\&\sqrt{1 + 1 + 0} \\&\sqrt{2} = 1.414\end{aligned}$$

Ken:

$$\begin{aligned}&\sqrt{(1-3)^2 + (2-2)^2 + (0-1)^2} \\&\sqrt{(-2)^2 + (0)^2 + (-1)^2} \\&\sqrt{4 + 0 + 1} \\&\sqrt{5} = 2.236\end{aligned}$$

Lily:

$$\begin{aligned}&\sqrt{(5-3)^2 + (1-2)^2 + (1-1)^2} \\&\sqrt{(2)^2 + (-1)^2 + (0)^2} \\&\sqrt{4 + 0 + 1} \\&\sqrt{5} = 2.236\end{aligned}$$

Audrey:

$$\begin{aligned}&\sqrt{(0-3)^2 + (1-2)^2 + (1-1)^2} \\&\sqrt{(-3)^2 + (-1)^2 + (0)^2} \\&\sqrt{9 + 1 + 0} \\&\sqrt{10} = 3.162\end{aligned}$$

Hazel:

$$\begin{aligned}&\sqrt{(1-3)^2 + (2-2)^2 + (1-1)^2} \\&\sqrt{(-2)^2 + (0)^2 + (0)^2} \\&\sqrt{4 + 0 + 0} \\&\sqrt{4} = 2\end{aligned}$$

Evelyn:

$$\begin{aligned}&\sqrt{(1-3)^2 + (2-2)^2 + (0-1)^2} \\&\sqrt{(-2)^2 + (0)^2 + (-1)^2} \\&\sqrt{4 + 0 + 1} \\&\sqrt{5} = 2.236\end{aligned}$$

Produced the following values:

'Scarlett': 1.414,
'Hazel': 2,
'Tony': 2.236,
'Mary': 2.236,
'Ken': 2.236,
'Lily': 2.236,
'Evelyn': 2.236,
'James': 2.449,
'Audrey': 3.162,
'Mason': 3.317

Based on 3NN, 7 Jane appears to like Pop music.

TASK 2

$$P(22 \mid \text{Rock}) = 1/1 = 1$$

$$P(22 \mid \text{Jazz}) = 0/1 = 0$$

$$P(22 \mid \text{Pop}) = 0/1 = 0$$

$$P(\text{Spain} \mid \text{Rock}) = 0/3 = 0$$

$$P(\text{Spain} \mid \text{Jazz}) = 0/3 = 0$$

$$P(\text{Spain} \mid \text{Pop}) = 3/4 = 0.75$$

$$P(F \mid \text{Rock}) = 1/5 = 0.2$$

$$P(F \mid \text{Jazz}) = 2/5 = 0.4$$

$$P(F \mid \text{Pop}) = 2/5 = 0.4$$

$$P(22) * P(F) * P(\text{Spain}) = (1/10 * 5/10 * 3/10)$$

$$15/1000 = 0.015$$

$$P(\text{Rock}) = 3/10 = 0.3$$

$$P(\text{Jazz}) = 3/10 = 0.3$$

$$P(\text{Pop}) = 4/10 = 0.4$$

$$P(22, F, \text{Spain} \mid \text{Rock}) = P(\text{Rock} \mid 22, F, \text{Spain}) *$$

$$P(22, F, \text{Spain}) / P(\text{Rock})$$

$$0.005 * 0.015 / 0.3$$

$$\underline{0.000066}$$

$$P(22, F, \text{Spain} \mid \text{Jazz}) = P(\text{Jazz} \mid 22, F, \text{Spain}) *$$

$$P(22, F, \text{Spain}) / P(\text{Jazz})$$

$$0.0005 * 0.015 / 0.3$$

$$\underline{0.0000525}$$

$$P(22, F, \text{Spain} \mid \text{Pop}) = P(\text{Pop} \mid 22, F, \text{Spain}) *$$

$$P(22, F, \text{Spain}) / P(\text{Pop})$$

$$0.00132 * 0.015 / 0.3$$

$$\underline{0.000066}$$

With normalization:

$$(1+1/9+11) * (1+1/9+11) * (0+1/9+11)$$

$$0.1 * 0.1 * 0.05$$

$$0.005$$

$$(0+1/9+11) * (3+1/9+11) * (0+1/9+11)$$

$$0.05 * 0.2 * 0.05$$

$$0.0005$$

$$(0+1/12+11) * (3+1/12+11) * (3+1/12+11)$$

$$0.0435 * 0.174 * 0.174$$

$$0.00132$$

Based on the conditional probability, Rock and Pop appear to be equally likely outcomes.