INTRODUCTION MACHINE LEARNING

EXERCISE 6

TEACHER:
Johannes Kiesel

GROUP:
Group 16

SUBMITTED BY:

Aaron Perez Herrera
Cesar Fernando Gamba Tiusaba
Chun Ting Lin
Olubunmi Emmanuel Ogunleye

Exercise 1 : Probability Basics (1 Points)

Which of the following statements are true?

- According to the Kolmogorov axioms the statement P(A) - P(A) = 0 holds.
- A function that fulfills the Kolmogorov axioms is a probability measure.
- Two events are statistically independent P(An B) = P(A) + P(B).
- Each subset A of a sample space 12 is an event.

The true statements are.

- A function that fulfills the Kolmogorov axioms is a probability measure. (true)
- Each subset A of a sample space 1 is an event. (true)

Exercise 2 : Probability Basics (Kolmogorov) (0 Points)

Prove the implications of the Kolmogorov axioms from the lecture (Theorem 7).

The conditions to prove are the following ones:

a) $Prove\ P(\emptyset) = 0$
$Using\ axiom\ III \dashrightarrow\ A = \emptyset\ and\ B = \ \emptyset$
$Se\ P(A\ U\ B) = \ \emptyset\ U\ \emptyset = 0 \dashrightarrow \quad P(\emptyset) = 0$

b) $Prove\ P(A') = 1 - P(A)$
$Using\ combination\ of\ Axiom\ II\ and\ III$
$P(all) = 1\ (Axiom\ 2) \rightarrow P(all) = P(A') + P(A) = 1\ (Axiom\ 3)$
$1 = P(A') + P(A) \rightarrow \quad P(A') = 1 - P(A)$

c) $Prove\ A\ \in B \rightarrow P(A) \leq P(B) \rightarrow \mathrm{P(B \setminus A) = P(B) - P(A)}$
$\mathrm{P(B) = P(A) + P(B \setminus A)} \rightarrow (Using\ Axiom\ III)$
$\mathrm{P(B \setminus A) = P(B) - P(A)}$

d) $Prove\ P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$\mathrm{P(A \cup B) = P(A) + P(B \setminus A)} \rightarrow Then\ substitue\ (Prove\ 3)$
$\mathrm{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$

Exercise 3 : Bayes' Rule (2+3=5 Points)

A hospital database contains diagnoses (C1 ... C5) for 8 patients along with binary observations of symptoms S1 ... S9:

Patient	Diagnosis	Symptoms								
		S1	S2	S3	S4	S5	S6	S7	S8	S9
1	C1	1	0	1	0	1	0	0	0	0
2	C2	0	1	0	1	1	0	1	0	0
3	C3	1	0	1	0	0	1	0	1	0
4	C4	0	1	0	1	1	0	1	0	0
5	C3	1	0	1	0	0	0	0	1	0

6	**C5**	0	0	0	0	1	0	0	0	1
7	**C3**	1	0	1	0	0	1	0	0	0
8	**C2**	0	1	0	0	0	0	1	0	0

a) Compute based on the database the prior probabilities P(Ci) for each diagnosis.

The total number of patients is 8.

- .The number of patients diagnosed with C1 is 1, so P(C1) = 1/8 = 0.125.
- The number of patients diagnosed with C2 is 2, so P(C2) = 2/8 = 0.25.
- The number of patients diagnosed with C3 is 3, so P(C3) = 3/8 = 0.375.
- The number of patients diagnosed with C4 is 1, so P(C4) = 1/8 =0.125.
- The number of patients diagnosed with C5 is 1, so P(C5) = 1/8 = 0.125.

b) Compute based on the database the posterior probabilities P(Ci | S4) for each diagnosis.

Posterior probabilities P(Ci|S4) of the diagnoses Ci given symptom S4 are calculated as follows:

The posterior probability of Ai given B is given by,

$$P(A_i|B) = \frac{P(A_i) * P(B|A_i)}{\sum_{i=1}^{k} P(A_i) * P\,(B|A_i)}$$

Hence, first P(B|Ai) is to be calculated.

$$P(B|A) = \frac{P\,(A \cap B)}{P(B)}, \qquad if\ P(B) > 0$$

Therefore, the posterior probability of S4 given C is calculated as:

$$P(S_4|C_i) = \frac{P(S_4 \cap C_i)}{P(C_i)}$$

$P\,(S_4|C_1) = \frac{P(S_4 \cap C_1)}{P(C_1)} = \frac{0}{1/8} = 0$	$P\,(S_4|C_2) = \frac{P(S_4 \cap C_2)}{P(C_2)} = \frac{1/8}{2/8} = 0.5$
$P\,(S_4|C_3) = \frac{P(S_4 \cap C_3)}{P(C_3)} = \frac{0}{3/8} = 0$	$P\,(S_4|C_4) = \frac{P(S_4 \cap C_4)}{P(C_4)} = \frac{1/8}{1/8} = 1$
$P\,(S_4|C_5) = \frac{P(S_4 \cap C_5)}{P(C_5)} = \frac{0}{1/8} = 0$	

The next probabilities

$$P(C_i|S_4) = \frac{P(C_i) * P(S_4|C_i)}{\sum_{i=1}^{k} P(C_i) * P\,(S_4|C_i)}$$

$$\sum_{i=1}^{k} P(C_i) * P(S_4|C_i)$$

$$= P(C_1) * P(S_4|C_1) + P(C_2) * P(S_4|C_2) + P(C_3) * P(S_4|C_3) + P(C_4) * P(S_4|C_4) + P(C_5) * P(S_4|C_5)$$

$$= \frac{1}{8} * 0 + \frac{1}{4} * \frac{1}{2} + \frac{3}{8} * 0 + \frac{1}{8} * 1 + \frac{1}{8} * 0 = \frac{2}{8}$$

$$\sum_{i=1}^{k} P(C_i) * P(S_4 C_i) = \frac{1}{4}$$

$$P(C_1|S_4) = \frac{P(C_1) * P(S_4|C_1)}{\sum_{i=1}^{K} P(C_i) * P(S_4|C_i)} = \frac{\frac{1}{8} * 0}{\frac{1}{4}} = 0$$

$$P(C_2|S_4) = \frac{P(C_2) * P(S_4|C_2)}{\sum_{i=1}^{K} P(C_i) * P(S_4|C_i)} = \frac{\frac{1}{4} * \frac{1}{2}}{\frac{1}{4}} = 0.5$$

$$P(C_3|S_4) = \frac{P(C_3) * P(S_4|C_3)}{\sum_{i=1}^{K} P(C_i) * P(S_4|C_i)} = \frac{\frac{3}{8} * 0}{\frac{1}{4}} = 0$$

$$P(C_4|S_4) = \frac{P(C_4) * P(S_4|C_4)}{\sum_{i=1}^{K} P(C_i) * P(S_4|C_i)} = \frac{\frac{1}{8} * 1}{\frac{1}{4}} = 0.5$$

$$P(C_5|S_4) = \frac{P(C_5) * P(S_4|C_5)}{\sum_{i=1}^{K} P(C_i) * P(S_4|C_i)} = \frac{\frac{1}{8} * 0}{\frac{1}{4}} = 0$$

Exercise 4

Part (a): Calculate P(A), P(B), P(C), and P(D):

Each event corresponds to the probability of a specific color of balls (green, blue, yellow, or red) being

in one of the eight boxes. Using the image provided:

Green balls (a) appear in 3 boxes.

Blue balls (b) appear in 4 boxes.

Yellow balls (c) appear in 2 boxes.

Red balls (d) appear in 4 boxes.

The probabilities are calculated as:

P(A) = 3 / 8 = 0.375

P(B) = 4 / 8 = 0.5

P(C) = 2 / 8 = 0.25

P(D) = 4 / 8 = 0.5

Part (b): Calculate P(A and B), P(A and C), P(B and C), and P(B and D):

The joint probabilities are calculated by counting the boxes containing both colors:

P(A and B) = 1 / 8 = 0.125

P(A and C) = 1 / 8 = 0.125

P(B and C) = 1 / 8 = 0.125

P(B and D) = 2 / 8 = 0.25

Part (c): Check for statistical independence:

Pair (A, B):

P(A) = 0.375, P(B) = 0.5, P(A and B) = 0.125 P(A) × P(B) = 0.375 × 0.5 = 0.1875

Since P(A and B) = 0.125 != 0.1875, A and B are not independent.

Pair (A, C):

P(A) = 0.375, P(C) = 0.25, P(A and C) = 0.125 P(A) × P(C) = 0.375 × 0.25 = 0.09375

Since P(A and C) = 0.125 != 0.09375, A and C are not independent.

Pair (B, C):

P(B) = 0.5, P(C) = 0.25, P(B and C) = 0.125 P(B) × P(C) = 0.5 × 0.25 = 0.125

Since P(B and C) = 0.125 = 0.125, B and C are independent.

Pair (B, D):

P(B) = 0.5, P(D) = 0.5, P(B and D) = 0.25 P(B) × P(D) = 0.5 × 0.5 = 0.25

Since P(B and D) = 0.25 = 0.25, B and D are independent.

Part (d): Calculate P(A|C), P(B|C), and P(A and B|C):

Conditional probabilities are calculated using the formula:

P(A|C) = P(A and C) / P(C)

P(B|C) = P(B and C) / P(C)

P(A and B|C) = P(A and B and C) / P(C)

From the data:

P(A|C) = 0.125 / 0.25 = 0.5

P(B|C) = 0.125 / 0.25 = 0.5

P(A and B|C) = 0 / 0.25 = 0

Part (e): Calculate P(B|D), P(C|D), and P(B and C|D):

Similarly, conditional probabilities given D are:

P(B|D) = P(B and D) / P(D)

P(C|D) = P(C and D) / P(D)

P(B and C|D) = P(B and C and D) / P(D)

From the data:

P(B|D) = 0.25 / 0.5 = 0.5

P(C|D) = 0 / 0.5 = 0

P(B and C|D) = 0 / 0.5 = 0

Part (f): Check for conditional independence:

Two events A and B are conditionally independent given C if:

$$P(A \text{ and } B \mid C) = P(A \mid C) \times P(B \mid C)$$

From (d), $P(A \text{ and } B \mid C) = 0 \mathrel{!=} 0.5 \times 0.5 = 0.25$ Thus, A and B are not conditionally independent given C

Exercise 5 : Naïve Bayes (3+2=5 Points)

Given is the following dataset to classify whether a dog is dangerous or well-behaved in character:

Color	**Fur**	**Size**	**Character** (C)
brown	ragged	small	well-behaved
black	ragged	big	dangerous
black	smooth	big	dangerous
black	curly	small	well-behaved
white	curly	small	well-behaved
white	smooth	small	dangerous
red	ragged	big	well-behaved

a) Determine the parameters P(Ai) and P(Bj=xj | Ai) for a Naïve Bayes classifier on this dataset. Use a table like the following to report the P(Bj=xj | Ai):

Prior Probability each Class

C1 (Well – Behaved)	= 4/7 = 0.57
C2 (Dangerous)	= 3/7 = 0.43

Conditional Probability

Attribute	Value (xj)	Class (ci)	P (xj | ci)
Color	Brown	Well-Behaved	1/1 = 1.00
Color	Brown	Dangerous	0/1 = 0.00
Color	Black	Well-Behaved	1/3 = 0.33
Color	Black	Dangerous	2/3 = 0.67
Color	White	Well-Behaved	1/2 = 0.50
Color	White	Dangerous	1/2 = 0.50
Color	Red	Well-Behaved	1/1 = 0.00
Color	Red	Dangerous	0/1 = 0.00
Fur	Ragged	Well-Behaved	2/3 = 0.67
Fur	Ragged	Dangerous	1/3 = 0.33
Fur	Smooth	Well-Behaved	0/2 = 0.00
Fur	Smooth	Dangerous	2/2 = 1.00
Fur	Curly	Well-Behaved	2/2 = 1.00
Fur	Curly	Dangerous	0/2 = 0.00
Size	Small	Well-Behaved	3/4 = 0.75
Size	Small	Dangerous	1/4 = 0.25
Size	Big	Well-Behaved	1/3 = 0.33
Size	Big	Dangerous	2/3 = 0.67

b) Classify the new example x = (black, ragged, small) using a Naïve Bayes classifier with the parameters you calculated in (a).

X = (Color = Black, Fur = Ragged, Size = Small) → Y(x) predicted?

$$P(C1|\,X) = P(C1) * P(Black|C1) * P(Ragged|C1) * P(Small|C1)$$
$$P(C1|\,X) = 0.57 * 0.33 * 0.67 * 0.75 = 0.095 \rightarrow 9.5\%$$

$$P(C2|\,X) = P(C2) * P(Black|C2) * P(Ragged|C2) * P(Small|C2)$$
$$P(C2|\,X) = 0.43 * 0.67 * 0.33 * 0.25 = 0.024 \rightarrow 2.4\%$$

$$Y(X) = argmax(P(C1|\,X), P(C2|\,X)) = \; P(C1|\,X) = 0.095 \rightarrow 9.5\%$$
$$[Assigned\ Class = Well - Behaved]$$

Exercise 6 : P Classification with Naïve Bayes (1+1+1+1+1+1=6 Points)

In this exercise, you will implement the Naïve Bayes classifier for predicting whether a given text was written by a human or generated by a language model. To make this task a bit easier, you will use a modified version of the dataset where all texts have been converted to Bag-of-words representations. As usual, there are test cases provided to check your implementation steps. Submit the file with your predictions for the test set along with your other solutions.

Download and use these files from Moodle:

- Text files for training, validation, and test sets:
  texts-train.tsv, texts-val.tsv, texts-test.tsv – These files contain the texts of the examples in the training, validation, and test sets, respectively. Use these files to extract features for the classification task.
- Bag-of-words features for training, validation, and test sets:
  bow-features-train.npy, bow-features-val.npy, bow-features-test.npy – The function to load these features is already implemented in the template.
- Labels for the training and validation sets:
  labels-train.tsv, labels-val.tsv – The function to load the classes is already implemented in the template.
- Template for the programming exercise:
  programming_exercise_statistical_learning.py. It contains function stubs for each function mentioned below. To run the program, use the following command (make sure the above-mentioned files are in the data folder):
  python3 programming_exercise_statistical_learning.py

a) Implement the function class_priors to estimate the prior probabilities P(Ai), where Ai is the event that an example has the class c, for all possible classes c occurring in a dataset D. The function receives an array of values c(x) for the x ∈ D as input, and returns a Python dictionary mapping the distinct classes c to their prior probabilities.

b) Implement the function conditional_probabilities to estimate the conditional probabilities P(Bj=xj |Ai), where Bj=xj is the event that the j'th feature has the value xj in an example x. The function receives two arrays as input, and returns a nested dictionary with the mapping class c→feature index j → feature value xj →probability P(Bj=xj |Ai). Consider using the class collections.defaultdict from the Python standard library to make this easier.

c) Implement the method fit of the class NaiveBayesClassifier using the functions implemented so far.

d) Implement the method predict of the class NaiveBayesClassifier, which takes as input a single feature vector x and returns the most probable class according to the Naïve Bayes model.
e) Implement the function train_and_predict, which shall fit a Naïve Bayes model on the training set, evaluate it on the validation set and return an array of predictions on test samples, similarly to previous exercises.
f) Implement the function extract_features to extract useful features from the provided texts for the classification task. Re-run the training and prediction steps using these new features. Aim to achieve the highest possible performance on the validation set by carefully selecting and engineering features that enhance the classifier's performance.

The solution to this part is attached with the Zip file.