

DM868/DM870/DS804: Data Mining and Machine Learning
Spring term 2023

Exercise 8: Bayes Optimal, Naive Bayes, Distributions

Exercise 8-1 Bayes Optimal (1 point)

We have a classification problem with two classes “+” and “−”, three trained classifiers h_1 , h_2 , and h_3 , with the following probabilities of the classifiers, given the training data D :

$$\Pr(h_1|D) = 0.5$$

$$\Pr(h_2|D) = 0.3$$

$$\Pr(h_3|D) = 0.2$$

For the three test instances o_1 , o_2 , o_3 , the classifiers give the following class probabilities:

$o_1 : \Pr(+ h_1) = 0.6$	$\Pr(- h_1) = 0.4$
$\Pr(+ h_2) = 0.2$	$\Pr(- h_2) = 0.8$
$\Pr(+ h_3) = 0.9$	$\Pr(- h_3) = 0.1$
$o_2 : \Pr(+ h_1) = 0.6$	$\Pr(- h_1) = 0.4$
$\Pr(+ h_2) = 0.6$	$\Pr(- h_2) = 0.4$
$\Pr(+ h_3) = 1$	$\Pr(- h_3) = 0$
$o_3 : \Pr(+ h_1) = 0.6$	$\Pr(- h_1) = 0.4$
$\Pr(+ h_2) = 0.6$	$\Pr(- h_2) = 0.4$
$\Pr(+ h_3) = 0$	$\Pr(- h_3) = 1$

We combine the three classifiers to get a Bayes optimal classifier. Which class probabilities will we get from this Bayes optimal classifier for the three test instances?

Exercise 8-2 Naïve Bayes (1 point)

The skiing season is open. To reliably decide when to go skiing and when not, you could use a classifier such as Naïve Bayes. The classifier will be trained with your observations from the last year. Your notes include the following attributes:

The weather: The attribute `weather` can have the following three values: `sunny`, `rainy`, and `snow`.

The snow level: The attribute `snow_level` can have the following two values: ≥ 50 (There are at least 50 cm of snow) and < 50 (There are less than 50 cm of snow).

Assume you went skiing 8 times during the previous year. Here is the table with your decisions:

weather	snow level	ski ?
sunny	< 50	no
rainy	< 50	no
rainy	≥ 50	no
snow	≥ 50	yes
snow	< 50	no
sunny	≥ 50	yes
snow	≥ 50	yes
rainy	< 50	yes

- Compute the *a priori* probabilities for both classes $\text{ski} = \text{yes}$ and $\text{ski} = \text{no}$ (on the training set)!
- Compute the distribution of the conditional probabilities for the two classes for each attribute.
- Decide for the following weather and snow conditions, whether to go skiing or not! Use the Naïve Bayes classifier as trained in the previous steps for your decision.

	weather	snow level
day A	sunny	≥ 50
day B	rainy	< 50
day C	snow	< 50

Exercise 8-3 Random Variables and Probability Distributions (1 point)

We played a lot with dice in the lecture. When we take the sum of n dice (a random variable), we get a probability distribution over the possible values. For just one die, this distribution is discrete with equal probabilities over $\{1, 2, 3, 4, 5, 6\}$. For two dies, the probabilities are unequally distributed over $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. How does the shape of the probability distribution develop with increasing n ?

Visualize the development of the shape with a little program or script.