Friedrich Greisiger (5722487)
Francisco Jose Cortes Aldaco (5978432)
Xenia Morera Martínez (5364350)
Lea Nickel (5153726)
Alexander Held (5943764)

Machine Learning

Assignment 01

Team: 15

30.10.2024

1 General questions

1. Stages: Pre-processing, Feature Extraction, Feature Selection, Model Building, Evaluation & Model Selection, Post-processing.

2. Supervised learning: the model is trained using labeled data, where each input comes with a corresponding output or label. The goal is for the model to learn a function that maps inputs to the correct outputs, allowing it to make accurate predictions on new, unseen data.

   In unsupervised learning, the model is trained on data without labeled outputs. The model's objective is to identify patterns, groupings, or structures within the data.

3. Overfitting: The model is too complex and learns noise in the training data. This will cause the model to perform well on training data but poorly on new unseen data.

   Underfitting: The model is too simple and fails to capture patterns. As a result it misses important relationships, which leads to poor performance on both training and new data.

2 Naive bayes

1. The probability that a car is stolen is $P(\text{yes}) = \frac{6}{10} = 0.6 = 60\%$.

   The probability that a car is not stolen is $P(\text{no}) = \frac{4}{10} = 0.4 = 40\%$.

   The probability that a car is red given it is stolen is $P(\text{red} \mid \text{yes}) = \frac{3}{6} = 0.5 = 50\%$.

   The probability that a car is a grand tourer given it is stolen is $P(\text{grand tourer} \mid \text{yes}) = \frac{2}{6} \approx 0.33 = 33\%$.

   The probability that a car is domestic given it is stolen is $P(\text{domestic} \mid \text{yes}) = \frac{2}{6} \approx 0.33 = 33\%$.

   The probability that a car is red given it is not stolen is $P(\text{red} \mid \text{no}) = \frac{1}{4} = 0.25 = 25\%$.

   The probability that a car is a grand tourer given it is not stolen is $P(\text{grand tourer} \mid \text{no}) = \frac{2}{4} = 0.5 = 50\%$.

   The probability that a car is domestic given it is not stolen is $P(\text{domestic} \mid \text{no}) = \frac{3}{4} = 0.75 = 75\%$.

2.
$$Z = \sum_{k=1}^{2} P(y = k) \prod_{i=1}^{3} P(x_i | y = k)$$

$$\begin{aligned} Z = {} & P(y = \text{yes}) \cdot P(x_1 = \text{red}|y = \text{yes}) \cdot P(x_2 = \text{grand tourer}|y = \text{yes}) \cdot P(x_3 = \text{domestic}|y = \text{yes}) \\ & + P(y = \text{no}) \cdot P(x_1 = \text{red}|y = \text{no}) \cdot P(x_2 = \text{grand tourer}|y = \text{no}) \cdot P(x_3 = \text{domestic}|y = \text{no}) \end{aligned}$$

$$Z = (0.6 \times 0.5 \times 0.33 \times 0.33) + (0.4 \times 0.25 \times 0.5 \times 0.75)$$

$$Z = 0.03267 + 0.0375 = 0.07017$$

$$P(y = \text{yes}|x_1 = \text{red}, x_2 = \text{grand tourer}, x_3 = \text{domestic}) = \frac{P(y = \text{yes}) \prod_{i=1}^{3} P(x_i|y = \text{yes})}{Z}$$

$$P(y = \text{yes}|x_1 = \text{red}, x_2 = \text{grand tourer}, x_3 = \text{domestic}) = \frac{0.6 \times 0.5 \times 0.33 \times 0.33}{0.07017}$$

$$P(y = \text{yes}|x_1 = \text{red}, x_2 = \text{grand tourer}, x_3 = \text{domestic}) = \frac{0.03267}{0.07017} = 46.67\%$$

3. Benefits: Naive Bayes is simple, computationally efficient, and effective for high-dimensional data like text. It performs well even with small datasets and can handle irrelevant features due to the independence assumption.

   Downsides: The strong independence assumption is often unrealistic, leading to poor performance with correlated features. Probability estimates may also be inaccurate, limiting its effectiveness in complex tasks.

4.

3 Ranking Losses

x	y	$\hat{y}^1$	$\hat{y}^2$
x_1	1	1	2
x_2	2	3	3
x_3	3	2	7

Tabelle 1: Table of the values given in the example $\hat{y}^1$ and $\hat{y}^2$

1. We can use the **pairwise ranking loss** function in order to evaluate how well the given models are performing.

   The formula of the pairwise ranking loss function for prediction
   haty is the following:

$$L(\hat{y}, y) = \sum_{i<j} \max(0, -(y_i - y_j) \cdot (\hat{y}_i - \hat{y}_j))$$

   True ranking values: $y = [1, 2, 3]$
   Ranking values model 1: $\hat{y}^1 = [1, 3, 2]$
   Ranking values model 2: $\hat{y}^2 = [2, 3, 7]$

   Evaluating the first model we obtain the following:

$$\begin{aligned} \text{when } (x_1, x_2) \rightarrow y_1 < y_2 \rightarrow y_1 - y_2 ß 1 - 2 = -1 \\ \hat{y}_1^1 < \hat{y}_2^1 \rightarrow \hat{y}_1^1 - \hat{y}_2^1 \rightarrow 1 - 3 = -2 \\ max(0, -(-1) \cdot (-2)) = max(0, -2) = 0 \end{aligned}$$

   Model result 1.1: Ranking is correct

$$\begin{aligned} \text{when } (x_1, x_3) \rightarrow y_1 < y_3 \rightarrow y_1 - y_3 ß 1 - 3 = -2 \\ \hat{y}_1^1 < \hat{y}_3^1 \rightarrow \hat{y}_1^1 - \hat{y}_3^1 \rightarrow 1 - 2 = -1 \\ max(0, -(-2) \cdot (-1)) = max(0, -2) = 0 \end{aligned}$$

   Model result 1.2: Ranking is correct

$$\begin{aligned} \text{when } (x_2, x_3) \rightarrow y_2 < y_3 \rightarrow y_2 - y_3 ß 2 - 3 = -1 \\ \hat{y}_2^1 < \hat{y}_3^1 \rightarrow \hat{y}_2^1 - \hat{y}_3^1 \rightarrow 3 - 2 = 1 \\ max(0, -(-1) \cdot (1)) = max(0, 1) = 0 \end{aligned}$$

Model result 1.3: Ranking is incorrect

Now valuating the second model we obtain the following:

$$\begin{aligned} \text{when } (x_1, x_2) \rightarrow y_1 < y_2 \rightarrow y_1 - y_2\text{ß}1 - 2 = -1 \\ \hat{y}_1^2 < \hat{y}_2^2 \rightarrow \hat{y}_1^2 - \hat{y}_2^2 \rightarrow 2 - 3 = -1 \\ max(0, -(-1) \cdot (-1)) = max(0, -1) = 0 \end{aligned}$$

Model result 2.1: Ranking is correct

$$\begin{aligned} \text{when } (x_1, x_3) \rightarrow y_1 < y_3 \rightarrow y_1 - y_3\text{ß}1 - 3 = -2 \\ \hat{y}_1^2 < \hat{y}_3^2 \rightarrow \hat{y}_1^2 - \hat{y}_3^2 \rightarrow 2 - 7 = -5 \\ max(0, -(-2) \cdot (-5)) = max(0, -5) = 0 \end{aligned}$$

Model result 2.2: Ranking is correct

$$\begin{aligned} \text{when } (x_2, x_3) \rightarrow y_2 < y_3 \rightarrow y_2 - y_3\text{ß}2 - 3 = -1 \\ \hat{y}_2^2 < \hat{y}_3^2 \rightarrow \hat{y}_2^2 - \hat{y}_3^2 \rightarrow 3 - 7 = -4 \\ max(0, -(-1) \cdot (-4)) = max(0, -4) = 0 \end{aligned}$$

Model result 2.3: Ranking is correct

Total loss of model 1: 1
Total loss of model 2: 0

(a) Conclusion: the second model is better at ranking since its total loss is 0

(b) The squared error is problematic in this case because it compares the predicted value against the true value instead of checking the ranking values.
In the examples given we can see that the first model's values $\hat{y}^1$ are closer to the true values y, while the values of the second model $\hat{y}^2$ are not that close, if we used the squared error it would suggest that the first model is performing better, but as it was shown in the previous exercise the first model presents some error while the second one is performing better at ranking.