Submission Date: 30th October 2024

1 General questions

To test your understanding of the first lecture, see if you can answer the following questions:

1. What are the stages of the ML cycle? Which ones are iterative stages?
2. What are the different types of learning?
3. How would you describe the overfitting and underfitting phenomenon?

2 Naive bayes

Naive bayes estimates the conditional probability of a label y given its features $X = (x_1, x_2, ..., x_n)$ as $P(y = k|X)$ where k is the class to be estimated. It is derived by making use of the chain rule of probabilities and the conditional independence assumption $P(x_i|x_{i+1}, ..., x_n, y) = P(x_i|y)$.

Naive bayes is defined as

$$P(y = k|x_1, x_2, ..., x_n) = \frac{1}{Z} P(y = k) \prod_{i=1}^{n} P(x_i|y = k) \tag{1}$$

where Z is a normalization constant called "evidence" that does not depend on y.

$$Z = \sum_{k=1}^{K} P(y = k) \prod_{i=1}^{n} P(x_i|y = k) \tag{2}$$

Given the following data:

Car	Color	Type	Origin	Stolen
1	red	sports	domestic	yes
2	red	sports	domestic	no
3	blue	sports	domestic	yes
4	blue	sports	domestic	no
5	blue	sports	imported	yes
6	blue	grand tourer	imported	no
7	blue	grand tourer	imported	yes
8	blue	grand tourer	domestic	no
9	red	grand tourer	imported	yes
10	red	sports	imported	yes

1. Estimate the probabilities by computing the relative frequencies: P(yes), P(red | yes), P(grand tourer | yes), P(domestic | yes), P(no), P(red | no), P(grand tourer | no), P(domestic | no)

2. Predict the probability that a car with properties x_1 = red, x_2 = grand tourer, x_3 = domestic will be stolen.

3. In general: What are the benefits, what are the downsides of using Naive bayes?

4. The extra mile: Derive Equation 1 using Bayes' theorem, the chain rule of probabilities and the conditional independence assumption stated above.

3 Ranking Losses

Assume that we want to evaluate a regression model that allows to rank inputs x. Therefore we are interested in having a prediction model that outputs a score (continuous value), which preserves the rank of the target, not necessarily the exact value of the target. For that, we are given two Machine Learning models with predicted values $\hat{y}_1$ and $\hat{y}_2$. The input data (x), target ranking or ground truth (y) and the output of the models are given in the following table:

x	y	$\hat{y}_1$	$\hat{y}_2$
x_1	1	1	2
x_2	2	3	3
x_3	3	2	7

1. Formulate mathematically a loss function that evaluates how well some generic prediction $\hat{y}$ matches the target ranking y.

2. According to this mathematical formulation, which model is better at ranking? Why is the squared error problematic in this case?

Hint: How could we reformulate this as a binary classification problem?