Home Assignment 1
Machine Learning A

NAME

September 7, 2022

1 Make Your Own (10 points)

1. Regarding the profile information, I think it would be interesting to find features which are correlated to the final grades.
These features could be:

- average grades in previous courses
- final grade in calculus
- final grade in statistics

So the sample space would be:

$$\mathcal{X} : \mathbb{R} \times \mathbb{N} \times \mathbb{N}$$

2. The label space would be the set of possible obtainable grades, considering the danish scale it is (maybe I'm wrong I don't know the danish scale yet, I'm sorry):
$\mathcal{Y} : \{12, 10, 7, 4, 2, 0\}$

3. The grades can be used as labels but also as numbers in order to perform regression, in case of regression, however when the algorithm will try to predict the results it will not obtain grades belonging to the $\mathbb{N}$ set, so I will have to map each $\mathbb{R}$ number predicted to the closest grade number after the prediction. For example 3,1 will become a 4.
Since I'm dealing with regression I can use the square-loss function:

$$l(y', y) = (y' - y)^2$$

to obtain grade that make sense, when I want to make a prediction the result has to be rounded to the closest grade
4. I would use the euclidean distance defined as:

$$d(x, x') = \sqrt{\sum_{i=1}^{d} (x_i - x'_i)^2}$$

but for KNN we could save computation just by doing

$$\sum_{i=1}^{d}(x_i - x_i')^2$$

5. I would divide the given dataset into: S_{train}, S_{val} and S_{test}. On S_{train} I would train the various instances of the model, we can use KNN as example.
on S_{val} I would take the best K; so the best possible classifier based on the previously defined error function. (the one with minimized loss function, so here I am evaluating the models based on the error function)
On S_{test} I evaluate the performance of the algorithm on new samples based on the same error function (this time unbiased). This last step will give me an idea of how the algorithm behave when I try to deploy it.

6. yes I would expect some issues, maybe the data on which I trained the algorithm wasn't very representative, so the deployed algorithm would fail to generalize, in this case, I would need more data in order to try to improve the algorithm performances.
It might also happen that the algorithm learns too well the noise and patterns between the data points and at the end will fail to generalize, in this case, I would try to use simpler models in order to obtain better results.

2 Digits Classification with K Nearest Neighbors (45 points)

3 Linear Regression (45 points)