Artificial Intelligence 2 – SS 2020
Assignment 13: Linear Regression in Python
– Given July 10., Due July 19. –

Hint: Exercises need to be handed in via StudOn at 23:59 on the day they are due or earlier. Please use only the exercise group of your tutor to hand in your work.

If any concepts here seem unfamiliar to you or you have no idea how to proceed, consult the lecture materials, ask a fellow student, your tutor, or on the Forum.

If a problem asks for code, comment it or make it otherwise self-explanatory. You do not need to write a lot, but it should be enough to convince your tutor that you understand what the code does. We may deduct up to 30% for uncommented and unclear code, but would prefer not to.

Problems with no points (0pt) will not be graded, but might appear on the exam in a similar form. For these, we will provide a reference solution after the submission deadline. If you find the reference solution unclear, ask about it on the forum or in in a tutorial.

Problem 13.1 (Titanic: Linear Regression in Python)

Implement *linear regression* in Python with `numpy`. You can find the necessary files at | 100pt
`https://kwarc.info/teaching/AI/resources/hw13-titanic.zip`.

This task is roughly based on the Titanic dataset on Kaggle[1]. It contains some information about passengers from the ship Titanic. Your task is to implement the functions `fit` and `predict` in `LinearRegression.py` which will predict passenger survival, as well as any additional data manipulation necessary in `TitanicSurvivalPrediction.py`. Which attributes are the strongest predictors of survival?

- In `LinearRegression.py` you will create your own linear regression predictor.
- In `TitanicSurvivalPrediction.py` you can implement any additional normalization and you need. Running this will produce a file `test_label.npy`, which we will use for grading.
- `train.npy` contains the training dataset, the labels for which you can find in `train_label.npy`. `test.npy` contains the test dataset.

Important: We will test your code automatically. So please make sure that:
- You use a recent Python version (≥ 3.5)
- You don't use any libraries (except for `numpy`)

Otherwise you risk getting no points.

You should upload `LinearRegression.py`, `TitanicSurvivalPrediction.py`, `test_label.npy`, and `answer.txt`. The last file should contain a list of the attributes that are the strongest predictors. To get this information you can train your classifier on a single attribute and compare the results. Finally justify whether your results meet your expectation!

---

[1]`https://www.kaggle.com/hesh97/titanicdataset-traincsv`