

Ethics for NLP: Spring 2022

Homework 0



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Due until Wednesday, 04.05.2022, 11:59pm

Submission Guidelines for Homework

- This homework worth 20 Points.
- Submit your code and answers in a single .ipynb notebook using Moodle.
- Name your submission file as: hw00.<matriculation_number>.lastname.firstname.ipynb
- No need to submit data files, since we already have it.
- Extra credit shall be given to well-structured submissions.
- In case of questions or remarks, please contact:
 - Aniket Pramanick, pramanick@ukp.informatik.tu-darmstadt.de

1 Classification (20 Points)

Before you start make sure you read the Submission Guideline instructions carefully associated to this assignment for important setup and submission information. Additionally, for this assignment you will be using **scikit-learn** library to code the machine learning models. For a brief tutorial you may follow [this link](#). For computing evaluation metrics use **sklearn** library.

1.1 XOR Function (6 Points)

Design and code a feed forward network to compute the function below, which is closely related to *XOR*:

$$f(x_1, x_2) = \begin{cases} -1 & \text{if } x_1 = 1 \text{ and } x_2 = 1 \\ 1 & \text{if } x_1 = 1 \text{ and } x_2 = 0 \\ 1 & \text{if } x_1 = 0 \text{ and } x_2 = 1 \\ -1 & \text{if } x_1 = 0 \text{ and } x_2 = 0 \end{cases} \quad (1)$$

The network should have a single output node that uses the “sigmoid” activation function:

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \quad (2)$$

Use a single hidden layer, with **ReLU** activation functions. Describe all weights and offsets.

1.2 Text Classification (14 Points)

After you extract the data from the **A0.tgz** file provided with this homework, in the directory *review_polarity*, you will find a dataset of positively and negatively classified reviews that was used by [this paper](#), a seminal paper about sentiment classification. Consult the **poldata.README.2.0** file inside the same folder for more information. Hold out a randomly selected 400 reviews as a test set.

1.2.1 Sentiment lexicon-based classifier: (7 Points)

Create a classifier using a sentiment lexicon. A lexicon from [this paper](#) is provided in the directory `opinion_lexicon_English`. Tokenize the data (use built-in Python function), and classify each document as positive if and only if it has more positive sentiment words than negative sentiment words (you need not perform any pre-processing steps like stop-words removal or deduplication of words).

Compute and report the accuracy and F_1 score (on detecting positive reviews) on the test set, using this lexicon-based classifier.

1.2.2 Logistic Regression Classifier: (7 Points)

Train a (binary) logistic regression classifier on your training set using features of your own choosing, and report its accuracy and F_1 score (as above) on the test set. In your write-up, describe the features you have chosen and explain the reasoning behind your choice.

Note: Do not use pre-trained word vectors or any features implemented or constructed by someone else.

For each of the following, write a review document that you believe would be considered as *positive* by human English speakers and:

- your lexicon classifier predicts it as *positive*, whereas your logistic regression classifier predicts it as *negative*.
- your lexicon classifier predicts it as *negative*, whereas your logistic regression classifier predicts it as *positive*.
- both of your classifiers predict it as *negative*.

For each of the above scenarios, briefly discuss why your classifier(s) would make incorrect predictions for the document you created.