MSDT: Normalization, regularization

Lisa Hensens

January 28, 2021

Terms

Normalization and regularization are different processes in machine learning and neural networks. With normalization, you adjust the data such as preprocessing and for us thus stemming for example. With regularization, you adjust the prediction function such as adding an extra part to the loss function.

What is already done?

In the algorithm that we have now, we already do the following things:

- Converting all letters to lower case
- Remove punctuation
- Stemming
- Tokenization
- Remove white space

Ideas for normalization

Inspiration for all these ideas has been taken from [1] [2].

- Converting numbers into words or removing numbers
- Removing stop words

Ideas for regulatization

- L2 & L1 regulatization
- Dropout

References

[1] https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization-techniques/

[2] https://medium.com/@datamonsters/text-preprocessing-in-python-steps-tools-and-examples-bf025f872908