

Machine Learning Revalida

Objectives

- To allow participants to apply their knowledge on Machine Learning in solving a real world problem
- To understand the technicalities of both the theoretical side of machine learning and application development in relation to it
- To be able to assess the uncertainty of machine learning results and how to move forward

Introduction

Natural Language Processing is a big problem in industry today as most of the data found in the real world are in the form of text. It has to be mined and understood before it can be translated into useful information for various stakeholders. Your objective for this project is to come up with a service that performs sentiment analysis that can be applied to 3rd party applications. You will incorporate what you learned in the machine / deep learning course using a specific medium -- text and determine if a submitted piece of text is considered one of several labels that you will define.

Instructions

1. Define the following terms in natural language processing. Make sure you can provide a simplified explanation as to how these techniques transform text into numbers and what it means in terms of developing your machine learning experiment:
 - a. Corpus
 - b. TF-IDF
 - c. Word Embeddings
 - d. Document Embeddings
 - e. Contextual Embeddings
2. Look for or implement your own transformations for items b to e in the previous instructions and make sure you can derive features from a given corpus.
3. Look for a dataset that relates to sentiment analysis and see if you can convert the text in the dataset into a set of feature vectors.
4. Create a deep learning model that classifies the selected dataset.
5. Incorporate your model as a serverless service using AWS Lambda that has the following properties:
 - a. Input: Some corpus text as { data: "your text here" }
 - b. Output: The classification of the text as { label: "happy" }

6. Create a client in Javascript or any other programming language that is able to communicate with the service in instruction item 5.
7. Be prepared with a simple presentation that incorporates the following:
 - a. Definition of Terms
 - b. Experimental methodology (i.e. how you created the machine learning experiment to determine what the best model / setup is)
 - c. Results of experimentation (i.e. what particular feature vector or model configuration yielded the best result? There should be a table of comparison of at least two configurations).
 - d. Demo of the end-to-end system (i.e. client interacting with the service)

List of Deliverables:

- Source code in github for both the service and the client
- Powerpoint slides

Rubric for Grading

Item	Percentage
Completeness of Work Requirements	50%
Explainability	30%
Analysis of Results	20%