CS165A (Spring 2023) Introduction to Artificial Intelligence

Discussion sections Week 1: April 5th, 2023

Overview

- Object-oriented ML design concepts
- Introduction to Project 1
- Concepts
  - Bag-of-words
- Demo:
  - Importance of Vectorization

OO ML design concepts

"classifer_agent" class in Project 1

- How to define such a class ?
- It's characteristics ?
- Do we need any methods ?
- Think in terms of python data structures, what can we use?

3 min Discussion!

“classifer_agent” class in Project 1

Attributes: (DS?)

- dictionary, vocabulary:
- parameters/weights:
- Feature map:

Methods: (i/o -> o/p)

- “predict”:
- “gradient”:
- “train”:
  - sgd/gd

Evaluation: (i/o -> o/p)

- “error”:
- “loss”:
- “logging”:

“classifer_agent” class in Project 1

Attributes: (DS?)

- dictionary, vocabulary: ?
- parameters/weights: dict/ndarray
- Feature map: (string → ndarray)
    - Instantiate object from dataproc class -> use anything
    - Initialize any “parsing” method as well

Methods: (i/o -> o/p)

- “predict”: string/ndarray→ which class? spam/non-spam
- “gradient”: labelled examples → gradient
- “train”: train data → wt. updates
    - SGD/GD

Evaluation: (i/o -> o/p)

- “error”: target data (→ error value: **R**
- “loss”: logistic/LS (ndarray) → value: **R**
- “logging”: saving/loading - helper functions

Bag-of-words features

- Simple and flexible way of extracting features from documents/textual data.
- A bag of words is a representation of text that describes the occurrence of words within a document. We just keep track of word counts and disregard the grammatical details and the word order.

“bag of words” Art in CMU MLD

Bag-of-words features

- It is called a “bag” of words because any information about the order or structure of words in the document is discarded.

- Only concerned with whether known words occur in the document, not where in the document.

Importance of Vectorization

https://colab.research.google.com/drive/12Ye7vYv3K4N-uPZDJylnLC7j6CDcRWq9?authuser=1#scrollTo=cexmNMmWnTR5

Q&A