

MA544_S21_PA2

May 31, 2021

0.1 # Programming Assignment - 2

0.2 ## Name:

0.3 ## Question 1

Modify the textual data example from class to convert the document-term matrix into TF-IDF (term-frequency inverse document frequency) matrix by using basic NumPy operations where

- Term-frequency (of a word in a document) is frequency of the word in a document divided by total number of words in the document.
- Inverse document frequency of a word (all documents under consideration) is the natural log of (total number of documents/number of documents having the given word).
- $\text{tf-idf}(t, d) = \text{tf}(t, d) * \text{idf}(t)$;

Display the dense output for the tf-idf matrix for the example taken in class.

```
[ ]: # Your code comes here

## Example Documents: Amazon review of Taco vs Burrito
document1 = " I bought this game as a gift for my 8 year old daughter who loves,
↳games. I was expecting lots of gross foods--but I was surprised at the,
↳inappropriate cards--eyeball, human burger, blood salsa, and fresh brains.
↳Those are not foods that typical people find in their refrigerators. We do,
↳not practice cannibalism. She was very upset when I suggested that we just,
↳take out those cards. I seriously wonder who thinks that those cards are,
↳appropriate for kids. The rest of the game is funny, but I wish I would have,
↳looked through the cards before I gave it to her."
document2 = "Absolutely love Taco vs burrito. I bought it as a kickstarter. I,
↳originally bought this game because my husband and I love to play games with,
↳friends but most of them are not targeted to children so I got this to add,
↳to our collection so we had options when our friends with kids came. I'm not,
↳gonna lie I did No have high expectations for this to be a game for adult,
↳but I was Sooooo wrong!!!!!! We have now played with several different,
↳groups of friends and it's a hit!!!!!! With adults it becomes a major,
↳strategy game. I have Now bought it as a Christmas present bc it was so well,
↳received!!!!"
```

```
document3 = " Unlike several of the reviewers here, I didn't purchase this_
↳originally for when kids are around. I bought it because of the reviews that_
↳said the adults all loved it too! I'm always on the lookout for games_
↳playable by 2 people and this was a great one. It's incredibly simple, but_
↳brings a lot of laughs with the competition and sabotage. I'm really glad I_
↳gave this game a chance."
doc_list = [document1, document2, document3]

## Create a bag of words

## Create a TF-IDF matrix
```