Plan tutorialu Tomek Korbak x Dawid Ratyński (26 listopada - 31 sierpnia)

Temat: Uczenie maszynowe w zastosowaniu do tekstu, obrazu i muzyki

Etapy:

1. Klasyfikacja tekstu (26 listopada - 16 lutego). Zadanie polega na przygotowaniu implementacji następujących modelu do prostego zadania klasyfikacyjnego (klasyfikacja języku dokumentu):
   1. Regułowy klasyfikator oparty o leksykon
   2. Naiwny klasyfikator Bayesowski
   3. Regresja logistyczna
   4. Lasy losowe

Efektem etapu będzie kod w postaci notebooków na platformie Google Colaboratory.

1. Klasyfikacja obrazu (17 lutego - 31 maja). Szczegóły do ustalenia.
2. Generowanie muzyki (1 czerwca - 31 sierpnia). Szczegóły do ustalenia.

# Materiały

<https://web.stanford.edu/~jurafsky/slp3/>

# 

# Notatki ze spotkań

## Spotkanie 26 listopada 2019

1. Znaleźć zbiory danych

2. Napisać pipeline do ewaluacji, accuracy

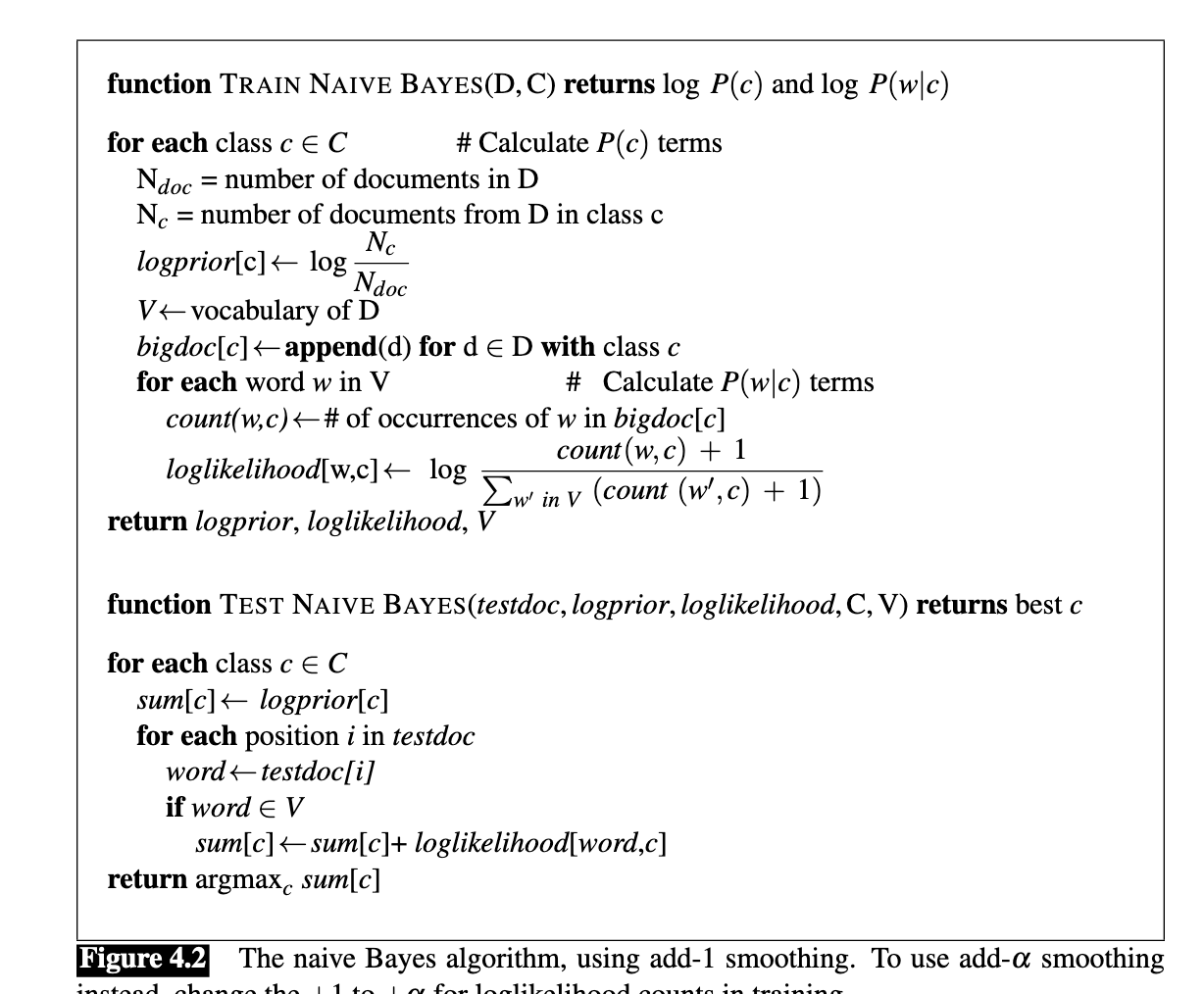
3. Bardzo prosty modelu opartego o leksykon (zbiory słów dla każdego języka)

4. Naive Bayes classifier - spróbować samemu, potem pobawić się implementacją z sklearna

P(słowa w dokuemencie|język) -> P(język|słowa w dokumencie)

https://scikit-learn.org/stable/modules/generated/sklearn.naive\_bayes.MultinomialNB.html#sklearn.naive\_bayes.MultinomialNB

https://web.stanford.edu/~jurafsky/slp3/4.pdf



https://scikit-learn.org/stable/modules/generated/sklearn.feature\_extraction.text.CountVectorizer.html

5. Regresja logistyczna

6. Lasy losowe

D = [0,0,0,0,0,0, 0, 2, 0] count vector