

## CSE 554/484

### Homework 02

Vektörleri oluşturdum en başta. Ve bu vektörleri Rocchio ve KNN algoritmaları için kullandım.

Class	Magazin	Siyasi	Spor	Sağlık	Ekonomi
Number	1	2	3	4	5

#### 1-)Rocchio

Prototype vektörü oluşturdum ve query vektörü ile  $\text{cosSim}(d, p_i)$  hesaplaması yaptım. En büyük  $\text{cosSim}$  değeri olan classification değerinin buldum. Similarity score ise  $\text{cosSim}$  değeridir.

File Name	12.txt	359.txt	663.txt	463.txt	276.txt
Number	5	4	3	2	1
Similarity Score	0.770433	0.513213	0.815603	0.698408	0.665677

#### 2-)K Nearest-Neighbour

Data set içerisindeki tüm dökümanların term weighted vektörleri ile query vektörünü  $\text{cosSim}$  işlemine soktum. Ardından bu değerleri sort ettim. Sonra sort edilmiş değerlere göre ilk k değerinin içerisinde en çok hangi classification var ise o classification değerinin döndürdüm.

File Name	KNN3	KNN5
12.txt	Class 5, similarity score = 1.0	Class 5, similarity score = 1.0
359.txt	Class 4, similarity score = 0.667	Class 1, similarity score = 0.6
663.txt	Class 3, similarity score = 0.667	Class 1, similarity score = 0.4
463.txt	Class 2, similarity score = 1.0	Class 2, similarity score = 1.0
276.txt	Class 1, similarity score = 0.667	Class 1, similarity score = 0.6

Ödevi 22 Mayıs 2017 23:45' te yükledim. Moodle'da 22 Mayıs 12pm gösterdiği için karışıklık oldu sanırım. Ödev geç yüklenmiş gözüküyor fakat ödevin pdf'inde 22 Mayıs gösteriyor, ben bu tarihe güvenip, bu tarihte yükledim.

Süleyman BALABAN  
121044014