Final Assignment – Spring 2023

Course Title: Machine Learning
Course Code: CSE 445
Date of Submission: 13 April 2023

---

1. Assume the following likelihoods for each word being part of a positive or negative movie review, and equal prior probabilities for each class.

	pos	neg
I	0.09	0.16
always	0.07	0.06
like	0.29	0.06
foreign	0.04	0.15
films	0.08	0.11

What class will Naive bayes assign to the sentence "I always like foreign films."?

2. Given the following short movie reviews, each labeled with a genre, either comedy or action:

1. fun, couple, love, love **comedy**
2. fast, furious, shoot **action**
3. couple, fly, fast, fun, fun **comedy**
4. furious, shoot, shoot, fun **action**
5. fly, fast, shoot, love **action**

and a new document **D: fast, couple, shoot, fly ?** compute the most likely class for D. Assume a naive Bayes classifier and use add-1 smoothing for the likelihoods.

3.

i. Find binary sentiment (class 1=positive or 0=negative) classification on the following movie review text.

ii. Derive cross-entropy loss for sentiment (class 1=positive or 0=negative) classification.

It's hokey. There are virtually no surprises , and the writing is second-rate . So why was it so enjoyable ? For one thing , the cast is great . Another nice touch is the music . I was overcome with the urge to get off the couch and start dancing . It upset me in , and it'll do the same to you . You may try it today.

Var	Definition	Value
x_1	count(positive lexicon) $\in$ doc)	3
x_2	count(negative lexicon) $\in$ doc)	2
x_3	$\begin{cases} 1 & \text{if "no"} \in \text{doc} \\ 0 & \text{otherwise} \end{cases}$	1
x_4	count(1st and 2nd pronouns $\in$ doc)	3
x_5	$\begin{cases} 1 & \text{if "!"} \in \text{doc} \\ 0 & \text{otherwise} \end{cases}$	0
x_6	log(word count of doc)	ln(72)

Suppose w = [-2.5, -5.0, -1.2, 1.5, 2.0, 1.7] b = 0.1

4. Cluster the following ten points (with (x, y) using k-means representing locations) into three clusters A1(2, 10) A2(2, 5) A3(8, 4) A4(5, 8) A5(7, 5) A6(6, 4) A7(1, 2) A8(4, 9) A9(9, 5), A10(1,4). Initial cluster centers are: A2(2, 5), A4(5, 8) and A7(1, 2).

5. Use the Nearest Neighbor clustering algorithm and Euclidean distance to cluster the examples from the previous exercise: A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9) A9=(9, 5), A10=(1,4). Suppose that the threshold t is 4.

6. Use the Maximin distance algorithm cluster the examples from the previous exercise: A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9) A9=(9, 5), A10=(1,4).