

	<b>Universidade Federal do Ceará</b>	
	<b>Disciplina:</b> TÓPICOS AVANÇADOS EM APRENDIZAGEM DE MÁQUINA / TÓPICOS ESPECIAIS EM LÓGICA I	<b>Código:</b> CK0255/CKP8366
	<b>Professor(a):</b> César Lincoln C. Mattos	
	<b>Semestre:</b> 2024.2	
	<b>Discente:</b>	<b>Matrícula:</b>
	<b>Curso:</b> Ciência da Computação	

## Trabalho 1

### Leia as Instruções:

- O trabalho é individual.
- As simulações poderão ser realizadas em quaisquer linguagens de programação.
- Para a avaliação do trabalho deverá ser submetido um arquivo pdf com texto e figuras referentes aos resultados das simulações.
- Para a avaliação do trabalho deverão ser enviados os códigos fonte.

1. Você joga uma moeda 10 vezes, observando 8 vezes o resultado cara ( $X = 1$ ). Considerando um modelo Beta-Bernoulli:

- Calcule a probabilidade  $\theta$  dessa moeda dar cara ( $X = 1$ ) considerando uma estimacão: (i) ML; (ii) MAP; (iii) média da posteriori (solução Bayesiana).
- Plote as distribuições (priori e posteriori) e a verossimilhança desse modelo considerando as priori  $p(\theta) = \text{Beta}(\theta|a = 1, b = 1)$  e  $p(\theta) = \text{Beta}(\theta|a = 2, b = 2)$ .

2. Implemente um modelo de Bernoulli Naive Bayes Bayesiano considerando os dados disponíveis em cleaned\_reviews.csv. Os dados consistem de reviews em texto (segunda coluna) da Amazon, classificados como positive, negative e neutral (primeira coluna). Ignore as demais colunas.

Use uma codificação bag-of-words, em que cada palavra é um atributo, com valor 0 (ausente) ou 1 (presente). Você pode usar o código abaixo:

```
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.preprocessing import LabelEncoder

data = pd.read_csv("cleaned_reviews.csv").dropna()
y = LabelEncoder().fit_transform(data.sentiments)
...
# split train-test
...
encoder = CountVectorizer(binary=True).fit(x_train)
x_train = encoder.transform(x_train)
x_test = encoder.transform(x_test)
```

Considerando 20 repetições aleatórias em que, para cara rodada, somente 1000 observações são usadas para treinamento e o restante para teste (escolhidos aleatoriamente):

- Mostre a média e o desvio padrão da acurácia do modelo no teste para a priori  $p(\theta) = \text{Beta}(\theta|a = 1, b = 1)$ , em que  $\theta$  é o parâmetro da verossimilhança de Bernoulli dos atributos.
- Repita o item anterior para as priori  $p(\theta) = \text{Beta}(\theta|a = 0.5, b = 0.5)$  e  $p(\theta) = \text{Beta}(\theta|a = 2, b = 2)$ .