

Bacharelado em Ciência da Computação - UFRPE/UAG
Reconhecimento de Padrões 2019.2
Prof. Dr. Luis Filipe

Lista de Exercícios 01

1. Acesse o endereço <http://wisdomquotes.com/> e baixe citações de 3 categorias distintas (ex.: *silence*, *funny*, *gratitude*). Escolha 150 citações de cada categoria.
2. Divida os dados em treinamento e teste.
3. Treine uma árvore de decisão com capacidade de classificar a categoria de novas citações. Você deve ter o acesso a regra utilizada para cada classificação.
4. Treine um classificador naive bayes com capacidade de classificar a categoria de novas citações. Você deve ter o acesso a probabilidade a posteriori para cada classe.
5. Compare os resultados dos classificadores estimando o erro real de cada um deles com 95% de confiança.
6. Para cada exemplo da base de teste, apresente: (a) a classe real, (b) a regra de classificação usada pela árvore; (c) a probabilidade a posteriori calculada pelo naive bayes.
7. Aumente o número de citações colhidas no site para o máximo possível. Para cada um dos classificadores, realize uma validação cruzada calculando a média e o desvio padrão das taxas de acerto obtidas. Compare com os resultados obtidos com os anteriores. Qual efeito observado ao aumentar as bases de dados?
8. Compare estatisticamente a performance dos dois classificadores através de um teste de hipótese.
9. Construa um sistema em que o usuário digita um texto e o programa o classifica em uma das categorias treinadas. O programa deve apresentar a regra de decisão utilizada, bem como a probabilidade a posteriori da classificação.

O seu código pode utilizar quaisquer recursos das bibliotecas sickit-learn e pandas.

O relatório deverá ser escrito utilizando o template “*Example of Elsevier article template with dummy text*” disponível no sistema overleaf. O documento deve conter:

- Título, Resumo, Introdução e Conclusão coerentes com um **trabalho de nível superior**;
- Descrição do método proposto contendo **TODAS** as justificativas para as decisões de projetos tomadas;
- Apresentação e explicação dos **trechos de códigos** mais significativos para a construção da solução;
- Detalhamento dos **resultados intermediários** obtidos ao longo da construção da solução.