

Machine Learning

CPS 863

Terceiro Trimestre de 2024

Professor: Edmundo de Souza e Silva

Lista de Exercícios 4

ATENÇÃO!

- Faça as listas de forma que TODAS AS RESPOSTAS sejam DEVIDAMENTE COMENTADAS (passos para se chegar a resposta).
- A entrega da lista deve ser feita em UM ÚNICO arquivo PDF. Não envie vários pedaços separadamente!
- ATENÇÃO! Faça as listas de forma que TODAS AS RESPOSTAS sejam DEVIDAMENTE COMENTADAS (passos para se chegar a resposta).

Não procure a solução na Internet ou em livros ou no chatGPT, pois o objetivo é que você mesmo avalie o que sabe. Obviamente, caso você já tenha conhecimento do problema, não leia a resposta (mesmo que já conheça o resultado final) e tente fazer sozinho. Só assim você poderá ter uma ideia melhor dos tópicos que você ainda não domina com desenvoltura.

- Anote as dúvidas encontradas para resolver **sozinho**. Em classe gostaria de saber quais as dúvidas que cada um teve para resolver o problema sem olhar a resposta.
- Qualquer referência a código é MUITO menos importante do que a EXPLICAÇÃO DOS PASSOS que foram realizados. O que mais importa é a explicação de como se chegou na solução.
- Para facilitar escrever a lista de forma clara, é possível traduzir equações a mão para LaTeX: <https://mathpix.com/>, ver também https://www.overleaf.com/learn/latex/Questions/Are_there_any_tools_to_help_transcribe_mathematical_formulae_into_LaTeX%3F

Questão 1

O Objetivo desta questão é adquirir experiência com o algoritmo *Expectation Maximization* (EM). Para isso você deve **implementar a sua solução**, que deverá estar bem explicada. (Não entre em pânico, a implementação é simples, e pode ser feita facilmente em Octave, R, Python, etc.)

O dataset fornecido contém o resultado de uma votação de mil usuários em 5 tipos de filmes. (Ao invés de filmes poderíamos ter candidatos a uma eleição ou outras *features*. O significado das *features* depende do problema, obviamente. Cada um dos usuários vota que acordo com o seu interesse um tipo de filme, por exemplo, filmes de Ação, Comédia, Romance, Suspense e Ficção Científica. Cada usuário dá uma nota de 1 a 4 para cada tipo de filme (ou *feature*) onde 1 significa “não gosto” e 4 “gosto muito”.

O objetivo é verificar se é possível classificar os usuários usando 2 classes. Para isso você escolhe o modelo *Naive Bayes* pela simplicidade. Deverá ser usado o algoritmo EM para essa tarefa.

- a) Quantos parâmetros deverão ser estimados para o modelo *Naive Bayes* a ser construído?
- b) Explique as equações usadas para resolver o problema.
- c) Baseado no item anterior, explique a sua implementação, incluindo as suas escolhas para a inicialização do código.
- d) Quantas iterações foram necessárias para resolver o problema? Qual o teste de parada utilizado?
- e) Quais os valores dos parâmetros encontrados? Quantos usuários foram alocados a cada uma das duas classes?

- f) O resultado da clusterização fez algum sentido? Explique e justifique a sua resposta.
- g) Qual a probabilidade do i -ésimo usuário ser alocado ao cluster 1? Explique sua resposta de forma genérica e escolha um dos 1000 usuários para exemplificar.
- h) Um usuário que votou 1,2,3,4,2 deveria ser alocado ao cluster 1 ou 2? Justifique.
- i) Como você classificaria um usuário que votou 3,2,?,2,3? Isto é, o usuário não votou na *feature* número 3 (por exemplo, não deu seu voto para filmes de Romance). Explique.

Questão 2 *Markov Chains*

Um robô pode se mover pelos quadrados da figura abaixo. A cada unidade de tempo (minuto, por exemplo) o robô tenta se mover para um dos quadrados adjacentes ao que ocupa e escolhe uma dentre as possíveis ações: (a) mover para Norte; (b) mover para Sul; (c) mover para Leste; (d) mover para Oeste. O robô **não** consegue enxergar a sua volta, então escolhe aleatoriamente uma dentre as 4 ações possíveis para só então fazer um movimento. Caso o robô tenha escolhido se mover na direção de um dos quadrados em rosa (os com tijolinhos na Figura 1) ou para fora da ambiente, ele colide com a parede e permanece na mesma posição até a nova tentativa.

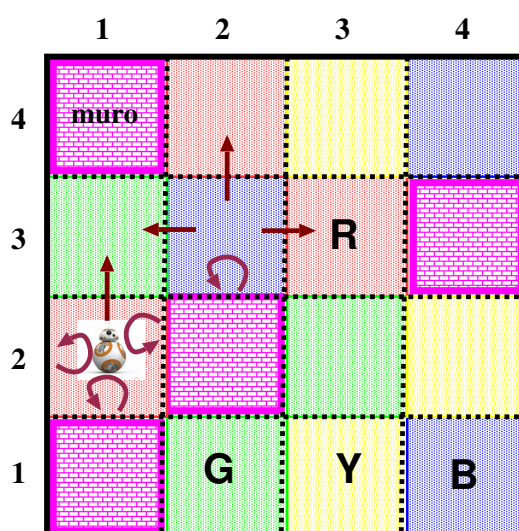


Figura 1: Robô andando por um ambiente

1. Construa a cadeia de Markov que representa o movimento do robô pelo ambiente, e mostre a matriz de transição de estados do sistema. Para facilitar a correção, indique claramente quais os estados do sistema e exemplifique indicando as probabilidades de transição para um movimento.
2. Suponha que o robô inicia a sua caminhada no quadrado indicado na figura (quadrado (2,1)). Qual a probabilidade do robô se encontrar no quadrado (1,3) vinte minutos após o início?
3. O robô está caminhando há muito tempo $t \rightarrow \infty$. Onde você apostaria que o robô se encontra?
4. Quantos minutos em média leva para o robô retornar ao ponto de partida?
É preciso pensar um pouquinho. Proponha uma possível solução a ser discutida na próxima aula. (Para a próxima aula, não precisa resolver esse item, mas será essencial mostrar que você pensou numa solução para apresentar)