

Objetivo

Este trabalho objetiva descobrir estruturas organizacionais em um conjunto de dados não marcados e exercitar conceitos de aprendizado de máquina não supervisionado.

Observações

Recomenda-se a utilização de técnicas de aprendizado não supervisionado tais como algoritmos de clusterização/agrupamento. As características representando cada texto são livres para cada indivíduo mas existem algumas leituras recomendadas na pasta do projeto. Pode haver discussões entre colegas mas cada pessoa precisa submeter sua própria solução e justificativas para suas escolhas. Lembre-se, isso pode ser visto como uma pequena competição.

Algumas perguntas que precisam ser respondidas no relatório

1. Quantos grupos existem nos dados?
2. Plote a divergência do agrupamento em função do número de grupos.
3. Onde você acha que o corte seria mais apropriado? Faça alguns experimentos neste sentido. Note que os dados podem conter desde 1 até 100 grupos. Cuidado com a escolha. Um ponto fundamental da avaliação será a descoberta do nro. de grupos e a forma como isso foi feito.
4. Mostre as mensagens "centróides" de cada grupo achado (esta parte pode ser colocada como anexo do relatório e pode conter até 5 páginas).
5. Mostre os 3 vizinhos mais próximos de cada centróide. Há sentido para todos os grupos?
6. Justifique cada escolha feita no relatório.
7. Um relatório final de até **5 páginas diretas e até 5 páginas de anexos**. O relatório deve ser feito explicando o método e os resultados obtidos.
8. A linguagem a ser utilizada é R.
9. O deadline é **Domingo 29/09 até 23:59min**.
10. Envie seu código final e relatório em arquivos separados intitulados
 - o mo444-trabalho-01-raXXXXXX.R
 - o mo444-trabalho-01-raXXXXXX.pdf
 - o Envie para o e-mail **anderson.rocha@ic.unicamp.br**
 - o **Assunto/subject:** "[MO444] Entrega Trabalho Individual #01".
 - o Sua mensagem será filtrada automaticamente se obedecer estes requisitos e você deverá receber uma mensagem de confirmação de submissão.

Dados

Os dados estão disponíveis em:

<http://www.ic.unicamp.br/~rocha/teaching/2013s2/mo444/trabalho-01/dados.tar.gz>

Valor

Este trabalho vale 10% da nota total.

Boa sorte.