



1. Ler o artigo

- Lozano, T. (2016) *Data Science: an Emerging Discipline*, CITAS.

Apresente três artigos, sendo um com ênfase em estatística, um com ênfase em ciência de dados e outro com ênfase em mineração de dados. Indique as diferenças e semelhanças nos dois estudos e pontos fortes e fracos de sua formação como estatístico para realizar estudos semelhantes.

Observações: Sua resposta deve ser resumida no máximo em uma página. Os estudos devem ser anexados a resposta. Este exercício pode ser realizado em grupos de até 5 alunos.

2. Pesquise sobre cada um dos problemas encontrados na análise multivariada. Para uma destas situações, encontre um exemplo de caso real e indique uma possível falha nas técnicas estatísticas tradicionais.

- (a) *Mining* (mineração, n muito elevado).
- (b) *Scalability* (escalabilidade).
- (c) *High Dimensional Data* (dados em alta dimensão, $n > p$).
- (d) *Pequenas amostras*.
- (e) *Curse of Dimensionality* (Problema de dimensionalidade).

3. Considere um hipercubo de dimensão r e lados de comprimento $2A$. Dentro deste hipercubo temos uma esfera r -dimensional de raio A . Encontre a proporção do volume do hipercubo que está fora da hipersfera e mostre que a proporção tende a 1 a medida que a dimensão r cresce. O que este resultado indica?

Para os exercícios seguintes, apresente o código R e interprete os resultados obtidos.

- 4. Johnson e Wichern - Exercício 1.1
- 5. Johnson e Wichern - Exercício 1.2
- 6. Johnson e Wichern - Exercício 1.6 - *air-pollution data*
- 7. Johnson e Wichern - Exercício 1.22 - *oxygen-consumption data*