

1 Parte I - Regressão Linear

O arquivo *casesBrazil.csv* contém dados do número oficial de casos do COVID-19 no Brasil, a partir de 25 de fevereiro de 2020, e se estendendo pelos 100 dias seguintes. A primeira coluna contém o índice do dia (dia 1, 2, 3, etc.) e a segunda contém o número de casos registrados.

1. Como vimos, a regressão linear pode ser usada também para o ajuste de curvas não lineares, bastando que se façam as transformações adequadas dos atributos. Com base nisso, implemente um algoritmo de regressão linear que ajuste um polinômio de grau n à curva do número de casos, usando **gradiente descendente** para a obtenção dos parâmetros. Faça testes com $n = 3, 5, 10$ e desenhe a curva de cada polinômio sobre os dados originais.
2. A inspeção visual dos dados mostra que a curva poderia ser melhor aproximada por uma curva exponencial, ou seja:

$$\theta_0 e^{\theta_1 x}.$$

Implemente um algoritmo de regressão linear com gradiente descendente que obtenha θ_0 e θ_1 . Mostre a curva exponencial obtida sobreposta aos dados originais. DICA: Em vez de aproximar y por $h_\theta(x)$, aproxime $\log y$ por $\log h_\theta(x)$.

3. Teste diferentes valores para a taxa de aprendizado α e mostre um gráfico do custo J em função de α . SUGESTÃO: Usar $\alpha = 10, 3, 1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001$.
4. Implemente o ajuste de $\theta_0 e^{\theta_1 x}$, mas agora usando equações normais.

2 Parte II - Regressão Logística

Um dos problemas de maior importância para o desenvolvimento histórico de *machine* e *deep learning* foi o reconhecimento de dígitos manuscritos. Neste contexto, a base de imagens MNIST¹ teve papel fundamental. A forma mais usual de reconhecer dígitos nessa base usando os algoritmos que aprendemos neste curso é linearizando cada imagem original de tamanho 28×28 para um vetor de comprimento 784. O arquivo *imageMNIST.csv* contém em cada linha uma imagem da base MNIST transformada em vetor. Já o arquivo *labelMNIST.csv* contém o dígito correspondente àquela linha.

- Implemente a regressão logística multi-classes para resolver este problema. Use vetorização sempre que possível. Mostre a porcentagem de imagens classificadas corretamente. Mostre também as imagens que ele classificou incorretamente, juntamente com o dígito verdadeiro e o dígito que o classificador atribuiu.
- Implemente agora a regressão logística multi-classes regularizada para o mesmo problema. Mostre a porcentagem de acerto e as imagens classificadas erroneamente.

¹<http://yann.lecun.com/exdb/mnist/>