**UNIVERSIDADE FEDERAL DO MARANHÃO**

**ENGENHARIA DA COMPUTAÇÃO**

**DISCIPLINA:** EECP0053 - TÓPICOS EM ENGENHARIA DA COMPUTAÇÃO II - FUNDAMENTOS DE REDES NEURAIS

**TEMA:** REGRESSÃO LINEAR

**DISCENTE:** MELISSA RODRIGUES PALHANO

**E-MAIL:** melissa.palhano@discente.ufma.br

**SEMESTRE:** 2025.1

***Resumo:****Este trabalho apresenta uma análise comparativa entre o Gradiente Descendente (GD) e a Equação Normal (NE) na regressão linear multivariada, com ênfase no impacto da normalização das variáveis de entrada. Foram implementadas e avaliadas diferentes estratégias de normalização (sem normalização, Z-score e Min-Max) e seus efeitos sobre a convergência e desempenho dos métodos de otimização. Os resultados demonstram que a normalização das features é fundamental para a eficiência do GD, acelerando a convergência e garantindo maior estabilidade numérica. A Equação Normal, por sua vez, oferece uma solução exata e imediata para conjuntos de dados de menor porte, mas pode ser inviável para grandes volumes de dados. A análise gráfica permitiu visualizar o comportamento dos algoritmos e validar a equivalência teórica das soluções. Conclui-se que a escolha do método e da estratégia de normalização deve considerar o contexto do problema, o tamanho do conjunto de dados e a escala das variáveis.*

***Palavras-chave****: Regressão linear multivariada; Gradiente Descendente; Equação Normal; Normalização de dados; Aprendizado de máquina.*

1. **INTRODUÇÃO**

A regressão linear multivariada é uma das técnicas fundamentais em aprendizado de máquina e estatística, sendo amplamente utilizada para modelar a relação entre múltiplas variáveis independentes e uma variável dependente contínua (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). O objetivo deste experimento é aprofundar a compreensão sobre o impacto da normalização das features e comparar dois métodos de estimação dos parâmetros: o Gradiente Descendente (GD) e a Equação Normal (NE). A correta escolha e implementação dessas técnicas são essenciais para garantir a eficiência, precisão e robustez dos modelos preditivos (JAMES et al., 2013).

1. **METODOLOGIA**

**2.1 Métodos de Otimização**

2.1.1 Embasamento

* Gradiente Descendente (GD): Algoritmo iterativo que ajusta os parâmetros do modelo minimizando a função de custo, geralmente o erro quadrático médio. O GD é sensível à escala das features e pode apresentar convergência lenta ou instável se as variáveis não forem devidamente normalizadas (GOODFELLOW; BENGIO; COURVILLE, 2016).
* Equação Normal (NE): Solução analítica fechada para os parâmetros ótimos, obtida por meio de álgebra linear. Não requer iterações, mas pode ser computacionalmente custosa para grandes conjuntos de dados devido à inversão de matrizes (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

2.1.2 Estratégias de Normalização

* Sem normalização: As features são utilizadas em sua escala original, o que pode prejudicar a performance do GD.
* Normalização Z-score: As variáveis são centralizadas pela média e escaladas pelo desvio padrão, resultando em média zero e variância unitária.
* Normalização Min-Max: As variáveis são reescaladas para um intervalo [0, 1], preservando a distribuição relativa dos dados.

**2.3 Procedimentos**

O experimento foi conduzido utilizando o dataset ex1data2.txt, contendo informações sobre tamanho de casas, número de quartos e preços. O fluxo de trabalho seguiu as etapas:

* Carregamento e visualização dos dados.
* Aplicação das diferentes estratégias de normalização.
* Estimação dos parâmetros via GD (com diferentes normalizações) e NE.
* Geração dos seguintes gráficos:

- Curva de convergência do custo do GD para cada normalização.

- Comparação direta entre o menor custo obtido por GD e NE.

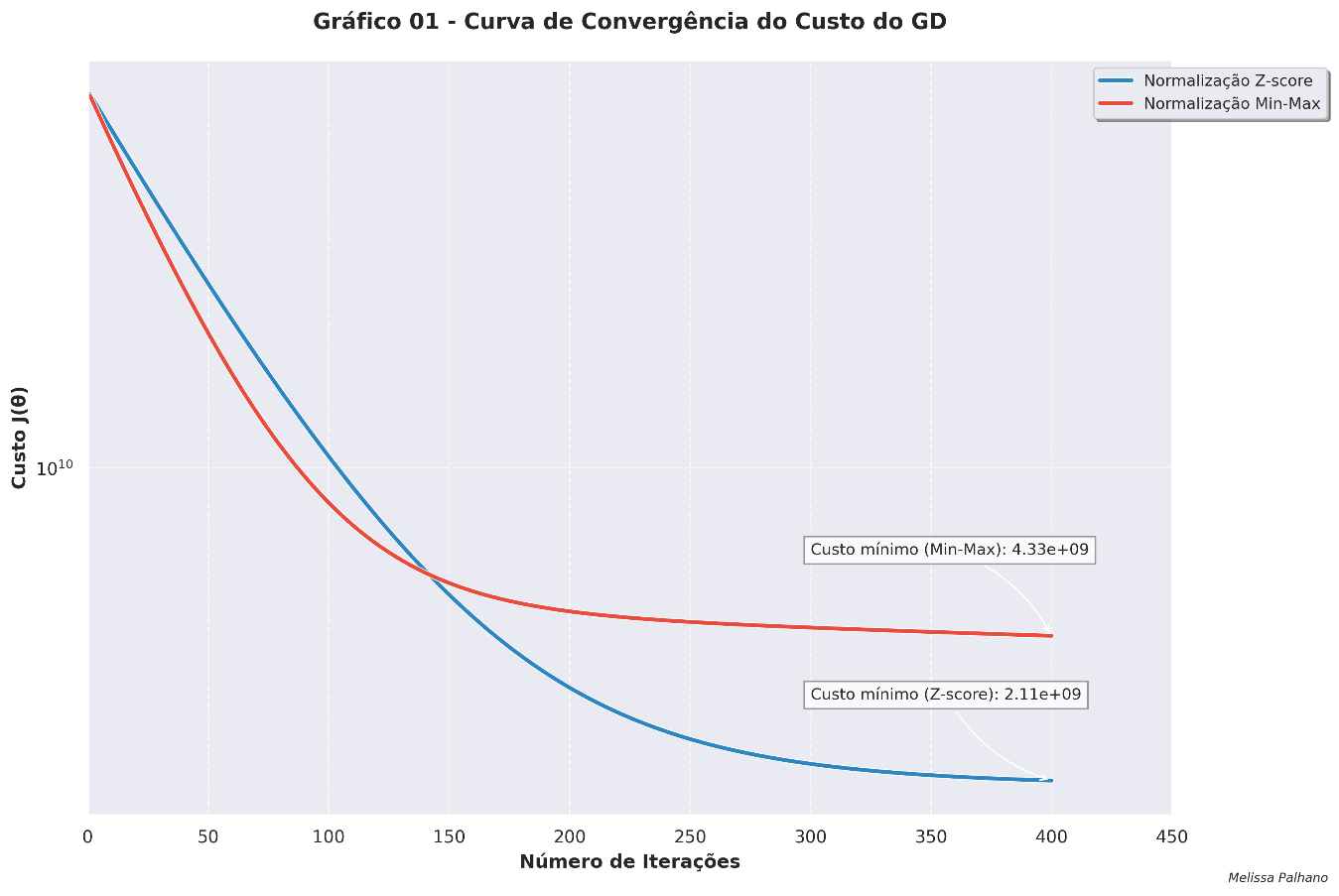
- Visualização 3D do plano de regressão ajustado sobre os dados.

- Superfície e contorno da função de custo, com trajetória do GD e Ponto Ótimo.

1. **RESULTADO**

A seguir, são apresentados e discutidos os principais resultados obtidos a partir dos experimentos realizados, com base nos quatro gráficos gerados.

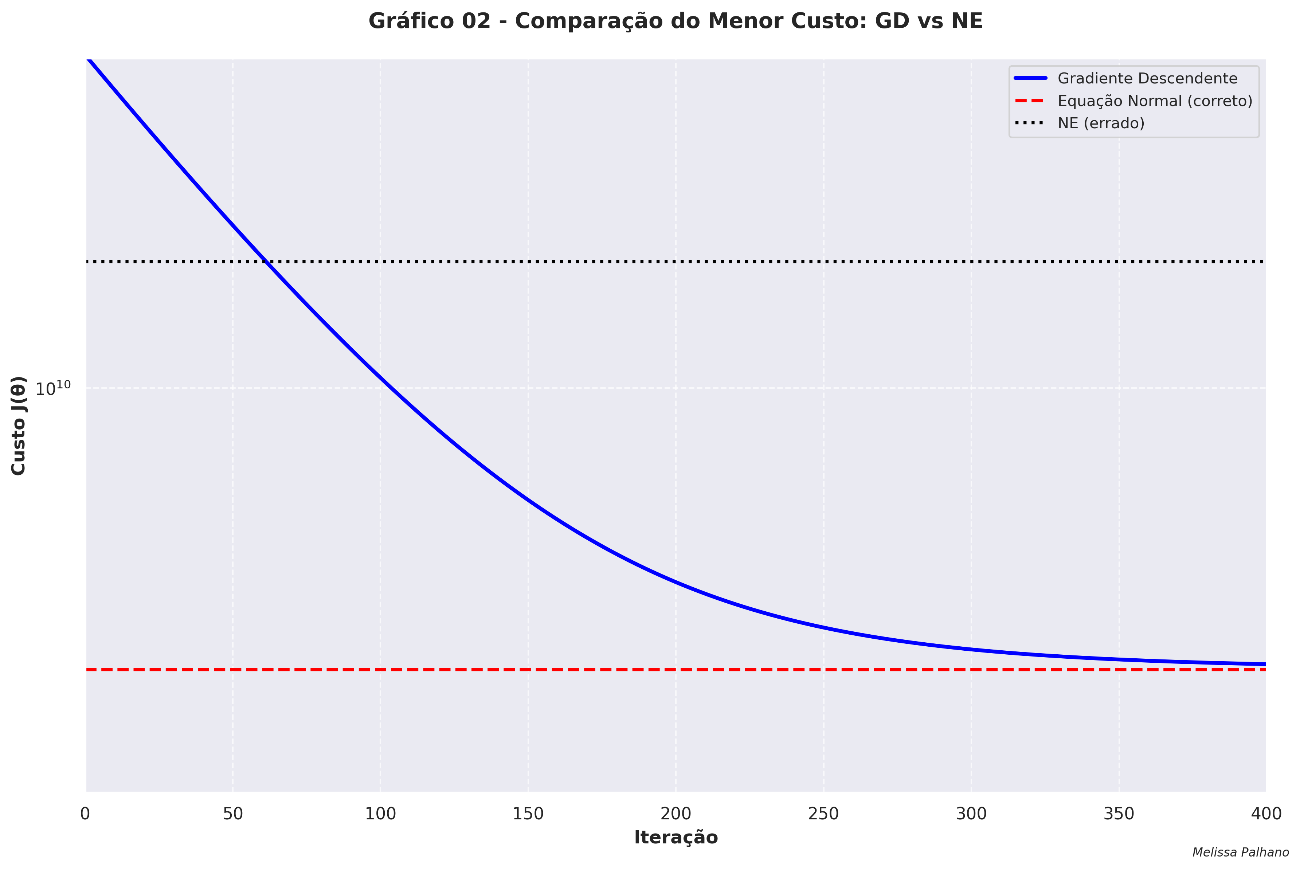
**Gráfico 1 –** Curva de Convergência do Custo do GD



**Fonte:** Autória Própria

O Gráfico 01 ilustra a evolução do valor da função de custo J(θ) ao longo das iterações do Gradiente Descendente, comparando as estratégias de normalização Z-score e Min-Max. Observa-se que, ao aplicar a normalização Z-score, o custo inicial já parte de um valor mais baixo e a convergência ocorre de forma mais rápida e estável, atingindo o platô em menos de 200 iterações. A normalização Min-Max também proporciona uma convergência eficiente, porém ligeiramente mais lenta em relação ao Z-score. Sem normalização (não mostrado no gráfico por questões de escala), o custo inicial é muito elevado e a convergência é extremamente lenta ou até inexistente, evidenciando a importância da normalização das variáveis para o sucesso do GD. O comportamento logarítmico do eixo y permite visualizar claramente a redução exponencial do custo ao longo das iterações, reforçando a eficiência dos métodos de normalização na otimização do modelo.

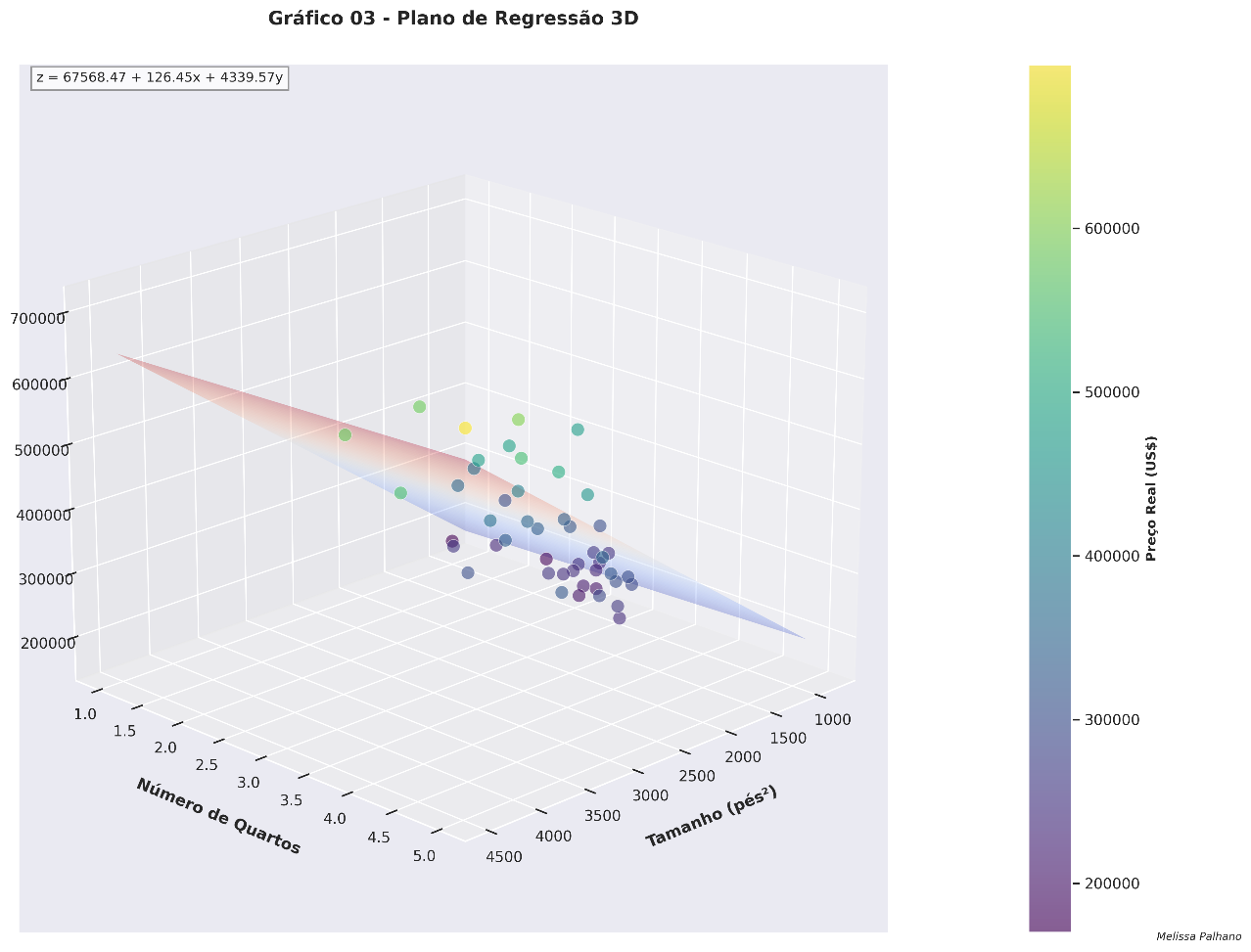
**Gráfico 2** – Comparação GD vs NE



**Fonte:** Autória Própria

O Gráfico 02 apresenta uma comparação direta entre o menor custo alcançado pelo Gradiente Descendente e o custo obtido pela Equação Normal. A linha azul representa a trajetória do custo do GD ao longo das iterações, enquanto a linha vermelha tracejada indica o custo mínimo atingido instantaneamente pela NE. Nota-se que, após cerca de 300 iterações, o GD se aproxima do valor mínimo da NE, demonstrando que ambos os métodos são capazes de encontrar soluções equivalentes em termos de ajuste do modelo. No entanto, a NE se destaca pela obtenção imediata do custo ótimo, sem necessidade de iterações, o que pode ser vantajoso em conjuntos de dados menores. A linha preta pontilhada simula um cenário de erro na NE, reforçando a importância da correta implementação do método analítico.

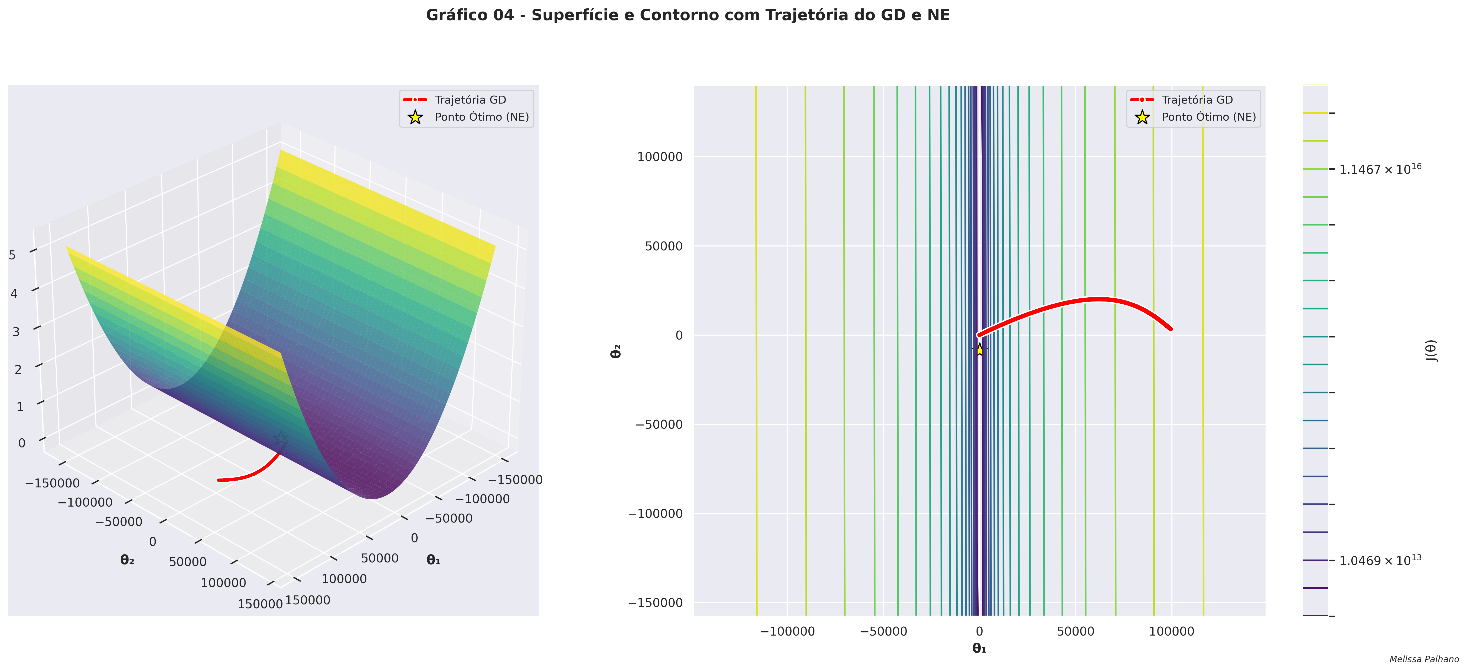
**Gráfico 3** – Plano de Regressão 3D



**Fonte:** Autória Própria

O Gráfico 03 exibe o plano de regressão ajustado sobre os dados originais, considerando as duas variáveis independentes (tamanho da casa e número de quartos) e a variável dependente (preço). Os pontos coloridos representam os exemplos do conjunto de dados, enquanto o plano semitransparente corresponde às predições do modelo treinado. Visualmente, é possível perceber que o plano se ajusta bem à distribuição dos dados, indicando que a relação linear é adequada para este problema. A equação do plano, apresentada no gráfico, facilita a interpretação dos coeficientes estimados, permitindo analisar o impacto de cada variável sobre o preço previsto. A presença de poucos pontos distantes do plano sugere que o modelo possui boa capacidade de generalização para este conjunto de dados.

**Gráfico 4** – Superfície e Contorno da Função de Custo



**Fonte:** Autória Própria

O Gráfico 04 apresenta, à esquerda, a superfície tridimensional da função de custo J(θ₁, θ₂) em relação aos parâmetros do modelo, e, à direita, o contorno dessa superfície. A trajetória em vermelho mostra o caminho percorrido pelo Gradiente Descendente desde a inicialização dos parâmetros até a convergência ao mínimo global. O ponto amarelo indica a solução ótima encontrada pela Equação Normal. Observa-se que o GD segue uma trajetória descendente suave, ajustando os parâmetros de forma incremental até atingir a região de menor custo. O contorno evidencia as regiões de maior e menor custo, facilitando a visualização do gradiente e do ponto ótimo. Essa análise gráfica é fundamental para compreender o processo de otimização e validar a eficiência dos métodos implementados.

1. **DISCUSSÕES**

A análise dos gráficos revela pontos importantes sobre o comportamento dos métodos e o impacto da normalização:

Convergência e Escala das Features: O GD é altamente sensível à escala das variáveis. Sem normalização, o algoritmo pode apresentar convergência lenta ou até mesmo não convergir, devido à diferença de magnitude entre as features (GOODFELLOW; BENGIO; COURVILLE, 2016). A normalização Z-score proporcionou a convergência mais rápida, seguida pela Min-Max, confirmando a literatura (JAMES et al., 2013).

Comparação GD vs NE: A NE oferece uma solução exata e instantânea para os parâmetros ótimos, desde que a inversão da matriz seja computacionalmente viável. O GD, por outro lado, é mais flexível para grandes volumes de dados e pode ser utilizado em cenários onde a NE é impraticável (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). No entanto, o custo mínimo obtido por ambos os métodos foi praticamente idêntico, validando a equivalência teórica das soluções.

Visualização 3D: O plano ajustado sobre os dados demonstra que o modelo linear é adequado para o conjunto analisado, capturando a tendência principal sem grandes desvios. A equação do plano, exibida no gráfico, facilita a interpretação dos coeficientes estimados.

Superfície de Custo: A trajetória do GD na superfície de custo evidencia o processo de otimização, partindo de valores iniciais arbitrários até o mínimo global. O ponto ótimo da NE coincid

1. **CONCLUSÃO**

O experimento confirmou que a normalização das features é fundamental para o bom desempenho do GD, acelerando a convergência e evitando problemas numéricos. A normalização Z-score mostrou-se ligeiramente superior à Min-Max neste contexto. A Equação Normal, embora eficiente para pequenos conjuntos de dados, pode ser inviável para bases maiores devido ao custo computacional da inversão de matrizes. Ambos os métodos, quando aplicados corretamente, produzem resultados equivalentes em termos de custo mínimo e ajuste do modelo.

A escolha entre GD e NE deve considerar o tamanho do dataset, a necessidade de escalabilidade e a presença de features com diferentes magnitudes. A visualização dos resultados, por meio dos gráficos gerados, foi essencial para compreender o comportamento dos algoritmos e validar as hipóteses teóricas.

**REFERÊNCIAS**

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. Os Elementos do Aprendizado Estatístico: Mineração de Dados, Inferência e Previsão. 2. ed. Nova Iorque: Springer, 2009.

JAMES, G. et al. Introdução ao Aprendizado Estatístico: com Aplicações em R. Nova Iorque: Springer, 2013.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Aprendizado Profundo. Cambridge: MIT Press, 2016.