www.datascienceacademy.com.br

Matemática Para Machine Learning

Como Usamos Derivadas em Machine Learning - Um Exemplo Prático

Cálculo é o estudo da relação entre variáveis e suas taxas de mudança. Utilizamos o Cálculo Diferencial como método para encontrar extremos de funções. Utilizamos o Cálculo Integral como método de modelagem probabilística.

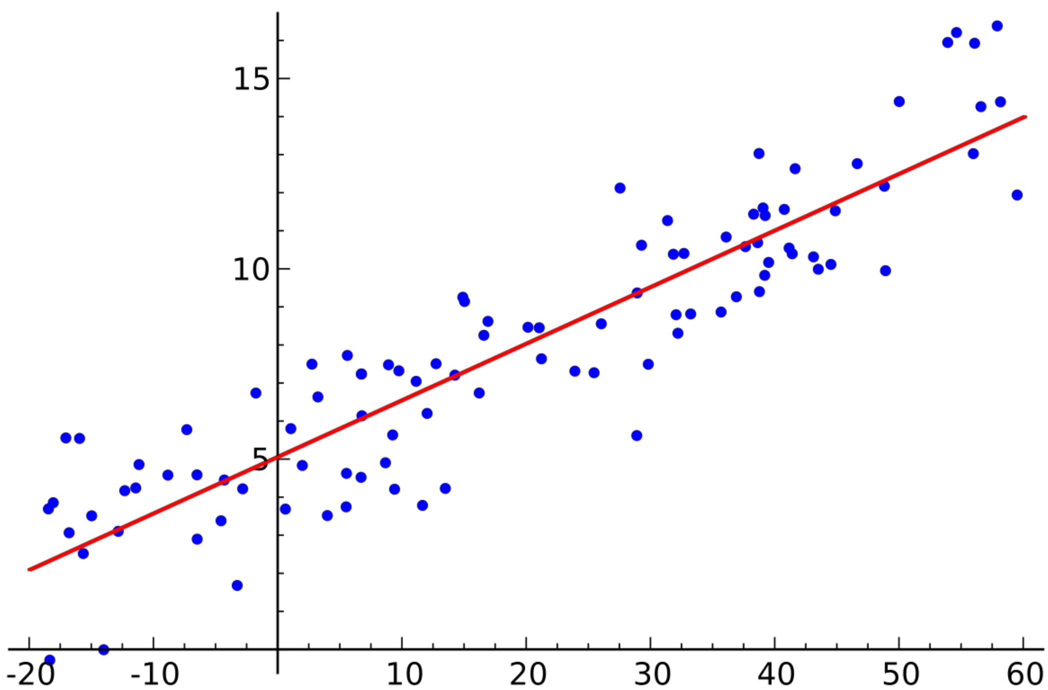
Derivadas (Cálculo Diferencial) são um conceito importante para aprendizado de máquina, assim como as Integrais (Cálculo Integral) são importantes em Estatística.

No processo de treinamento de muitos algoritmos de aprendizado de máquina, usamos gradiente descendente (ou um de seus parentes próximos). A descida de gradiente usa a derivada da soma de erros para atualizar um pouco os parâmetros do sistema de forma que o erro diminua o máximo possível.

Após cada atualização, o sistema aprende a prever com um erro menor. Deixe-o executar muitas iterações e ele convergirá em alguns ótimos (locais). O sistema está pronto para começar a fazer previsões.

O algoritmo de gradiente descendente é frequentemente usado como uma implementação vetorizada que usa cálculos de matrizes para cálculos altamente paralelos em uma GPU (Unidades de Processamento Gráfico).

Vejamos um exemplo concreto!!!



Um problema clássico de Machine Learning é a Regressão Linear. Suponha que tenhamos diversos pontos (x1, y1), (x2, y2),. . . , (xN, yN), e queremos encaixar uma linha da forma y = mx + b, como da figura acima.

Se eu tiver muitos pontos, é improvável que haja uma linha que passe exatamente por todos eles. Então, podemos pedir uma linha y = mx + b que fica o mais próximo possível dos pontos.

Uma opção fácil é usar o erro quadrado como medida de proximidade. Para um ponto (xn, yn) e uma linha definida por m e b, podemos medir o erro ao quadrado como:

$$[(mx_n + b) - y_n]^2$$

Ou seja: nosso **valor previsto menos o valor real**, **elevado ao quadrado**. Apenas para ilustrar: os valores reais (valores de y) são os valores dos pontos em azul no gráfico acima, enquanto os valores previstos (valores de mx + b) são os pontos na reta vermelha.

Podemos somar todos os erros pontuais para obter um erro total (o que vamos chamar de função de custo "J"):

$$J(m, b) = \sum_{n=1}^{N} \left[(mx_n + b) - y_n\right]^2$$

Note que escrevemos o erro J como uma função de m e b, pois, para qualquer configuração de m e b, obtemos um erro diferente (você lembra do conceito de função, certo?).

Agora, nosso objetivo é encontrar valores de m e b que minimizem o erro. Como podemos fazer isso? O Cálculo Diferencial nos diz que o mínimo da função J pode ser calculado encontrando os zeros de suas derivadas (supondo que seja convexo, conforme vimos nas aulas até aqui neste capítulo).

A derivada de uma função em um ponto é a inclinação da função naquele ponto (a derivada é como a velocidade). Para ser preciso, suponha que tenhamos uma função f que mapeie os números reais para números reais. (Isto é: f: R → R). Por exemplo,

$$f(x) = 3x^2 - e^x$$

A derivada de f com em relação a x, denotado ∂f / ∂x, é:

$$\frac{\partial f}{\partial x}(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

Isso essencialmente diz que a derivada de f em relação a x, avaliada em um ponto x0, é a taxa de mudança de f em x0. É bastante comum ver ∂f / ∂x denotado por f'.

Algo interessante sobre as derivadas é que elas nos permitem encontrar pontos extremos de funções de maneira direta (normalmente você pode pensar em um ponto extenso como um máximo ou mínimo de uma função). Podemos facilmente ver que o ponto em que a função é minimizada tem uma derivada (inclinação) de zero. Assim, se podemos encontrar zeros da derivada de uma função, também podemos encontrar mínimos (ou máximas) dessa função.

Como sempre dissemos: tudo se resume a um problema matemático!