

Trabalho de Complementos de Calculo - MATA95

Alunos: Raphael Matos da Costa, João Miguel Cândido Anastácio Costa

Tema: Derivadas Parciais

Definição:

Em matemática, derivadas parciais são uma extensão do conceito básico de derivadas, porém aplicada para funções de múltiplas variáveis. Cuja a variação da derivada é realizada em relação a uma das variáveis independente, mantendo as demais variáveis constantes. Este conceito é amplamente utilizado no cálculo vetorial, geometria diferencial e suas aplicações.

Definição formal:

Assim como em derivada simples, derivadas parciais são definidas como sendo um limite. Sendo U um conjunto aberto de \mathbb{R}^n e $f: U \rightarrow \mathbb{R}$ uma função. A derivada parcial de f no ponto $a = (a_1, \dots, a_n) \in U$ com relação à i -ésima variável a_i é definida por:

$$\frac{\partial}{\partial a_i} f(a) = \lim_{h \rightarrow 0} \frac{f(a_1, \dots, a_{i-1}, a_i + h, a_{i+1}, \dots, a_n) - f(a_1, \dots, a_i, \dots, a_n)}{h}$$

Mesmo se todas as derivadas parciais $\frac{\partial}{\partial a_i} f(a)$ existirem em um determinado ponto a , a função não necessita ser contínua.

Porém se todas as derivadas parciais existirem numa vizinhança de a e são contínuas, então f é o total derivado em que a derivada da vizinhança total é contínua.

Teorema de Schwarz:

Enunciado:

Seja f uma função de duas variáveis $f(x, y)$ definida em um conjunto aberto D . O teorema afirma que, se as derivadas parciais mistas de segunda ordem:

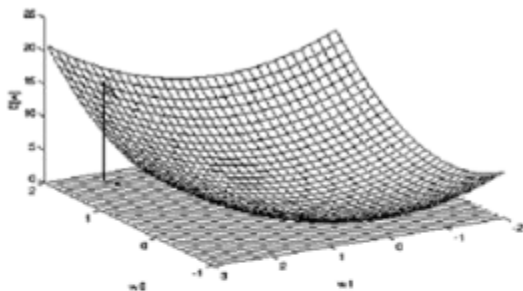
$$\frac{\partial^2 f}{\partial x \partial y} \text{ e } \frac{\partial^2 f}{\partial y \partial x}$$

Existem e são contínuas em um ponto (x_0, y_0) , ou em um disco aberto contido em D que contém esse ponto), então elas são iguais nesse ponto:

$$\frac{\partial^2 f}{\partial x \partial y} (x_0, y_0) = \frac{\partial^2 f}{\partial y \partial x} (x_0, y_0)$$

O que podemos interpretar desse teorema, a ordem de escolha das derivação em relação às diferentes variáveis que compõem a função indeferem, desde que as derivadas sejam contínuas.

Exemplos:



Gradiente

$$\nabla E[\vec{w}] = \left[\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \dots, \frac{\partial E}{\partial w_n} \right]$$

Regra de treinamento

$$\Delta \vec{w} = -\eta \nabla E[\vec{w}]$$

Ou seja,

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i}$$

A imagem anterior ilustra uma descida de gradiente em uma backpropagation de um rede neural artificial.

Motivação:

Escolhemos derivadas parciais pois a mesma é amplamente utilizada em rede neurais artificiais para definir o grau de enviesamento dentro do machine learning. Principalmente sendo utilizada na técnica de backpropagation. O backpropagation é aplicação eficiente da regra da cadeia de cálculo para calcular as derivadas parciais de forma iterativa, começando da camada de saída e movendo-se para trás (backward) através da rede. O objetivo é determinar o quanto a mudança em cada peso da rede afeta a perda (erro) total da rede. O mecanismo para isso é o algoritmo que propaga o erro de volta da camada de saída para as camadas anteriores, permitindo que a rede ajuste seus pesos durante o treinamento usando um otimizador como o Gradiente Descendente (Gradient Descent). O treinamento de uma rede neural visa minimizar a função de perda L ajustando os pesos w em cada camada. Para isso, precisamos saber a taxa de variação da perda em relação a cada peso, que é dada pela derivada parcial: $\frac{\partial L}{\partial w}$

Como a rede neural é uma composição de muitas funções (uma para cada neurônio e cada camada), a derivada de L em relação a um peso w de uma camada inicial deve levar em conta o efeito desse peso nas ativações subsequentes, até a saída final. O backpropagation usa a regra da cadeia para decompor esse cálculo em etapas localizadas e eficientes. Para calcular o gradiente é necessário o conjunto de todas essas derivadas parciais ($\frac{\partial L}{\partial w}$ para todos os w) forma o gradiente. O backpropagation calcula esse gradiente, que é então usado na etapa de otimização (por exemplo, $w_{\text{novo}} = w_{\text{antigo}} - \eta \frac{\partial L}{\partial w}$ para atualizar os pesos na direção que minimize a perda.