Gradient Descent

Formação Cientista de Dados

Machine Learning

- Loss Function/Cost Function: diferença entre a previsão e o valor real

Root Mean Squared Error (RMSE)

Independente de Escala

- O desvio padrão da amostra da diferença entre o previsto e o teste

Previsto	Realizado	Dif. ao Quad.
3,34	3,00	0,1156
4,18	4,00	0,0324
3,00	3,00	0
2,99	3,00	1E-04
4,51	4,50	1E-04
5,18	4,00	1,3924
8,18	4,50	13,5424

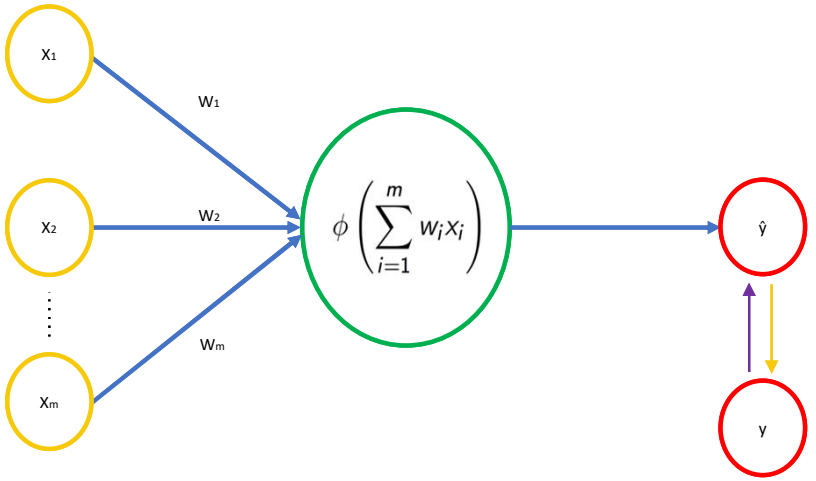
$$RMSE = \sqrt{\frac{\sum_{I=1}^{N}(p_i - t_i)^2}{N}}$$

$$RMSE = \sqrt{\frac{15{,}083}{7}}$$

$$RMSE = 1{,}46$$

Calculada a Loss Function...

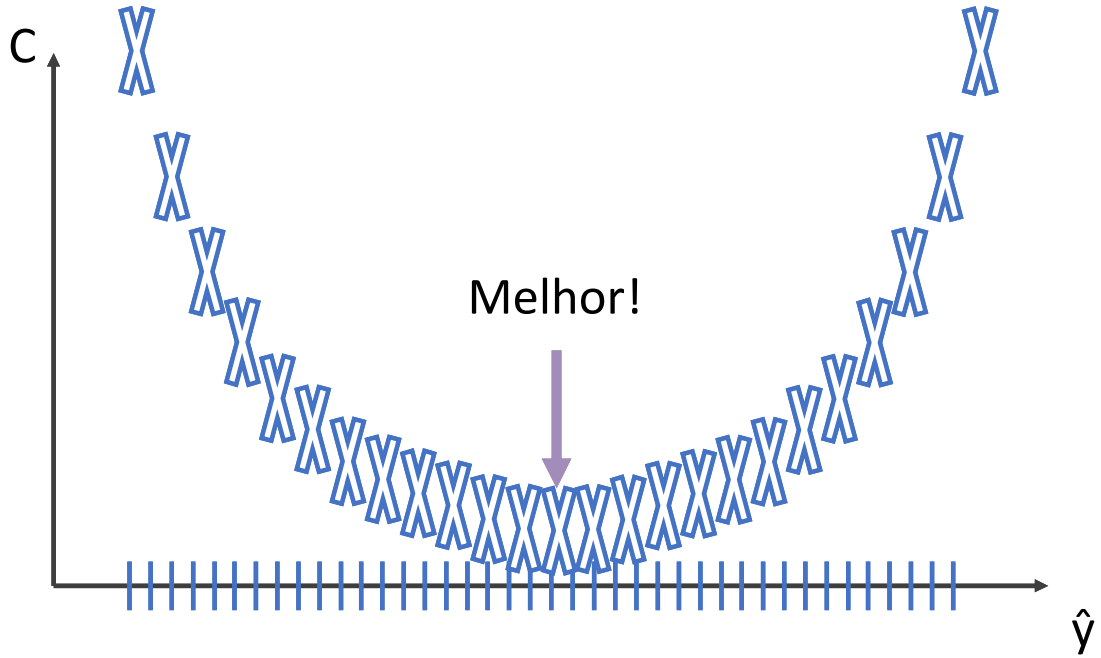
- É preciso atualizar os pesos da RNA...
- Backprogation



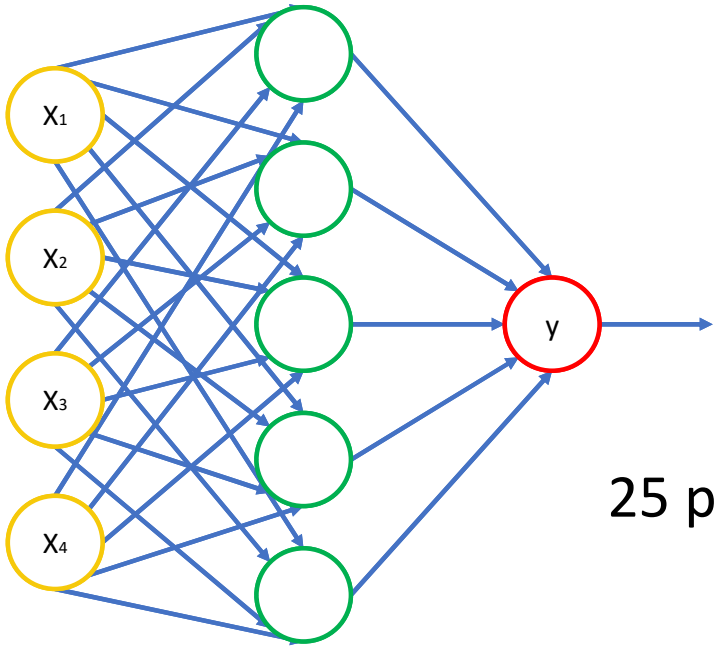
Ajustando os Pesos

- Learning Rate: taxa de aprendizado
- Como ajustar os pesos? Aumentar, diminuir? Em quanto?
- Lembrando que podemos ter muitos pesos para ajustar!

Força Bruta...



Gradient Descent



25 pesos

Gradient Descent

$1.000 \times 1.000 \times \ldots \times 1.000 = 1.000^{25} = 10^{75}$ combinações

Sunway TaihuLight: o super computador mais rápido do mundo

93 PFLOPS

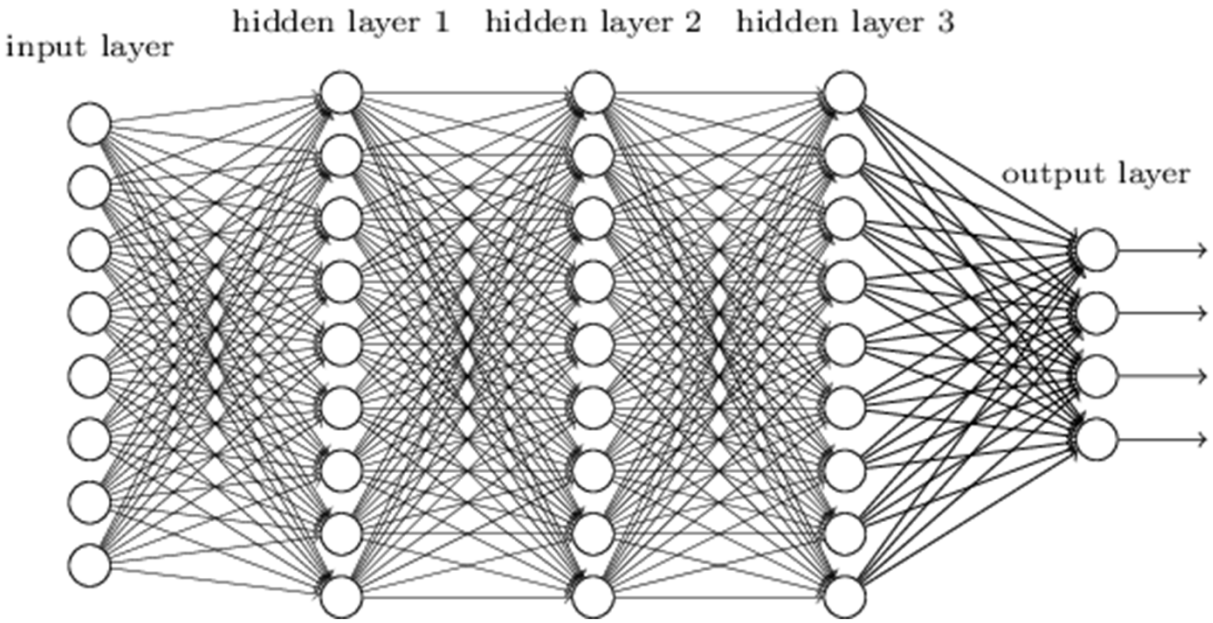
93×10^{15}

$10^{75} / (93 \times 10^{15})$

$= 1{,}08 \times 10^{58.}$ segundos

$= 3{,}42 \times 10^{50.}$ anos

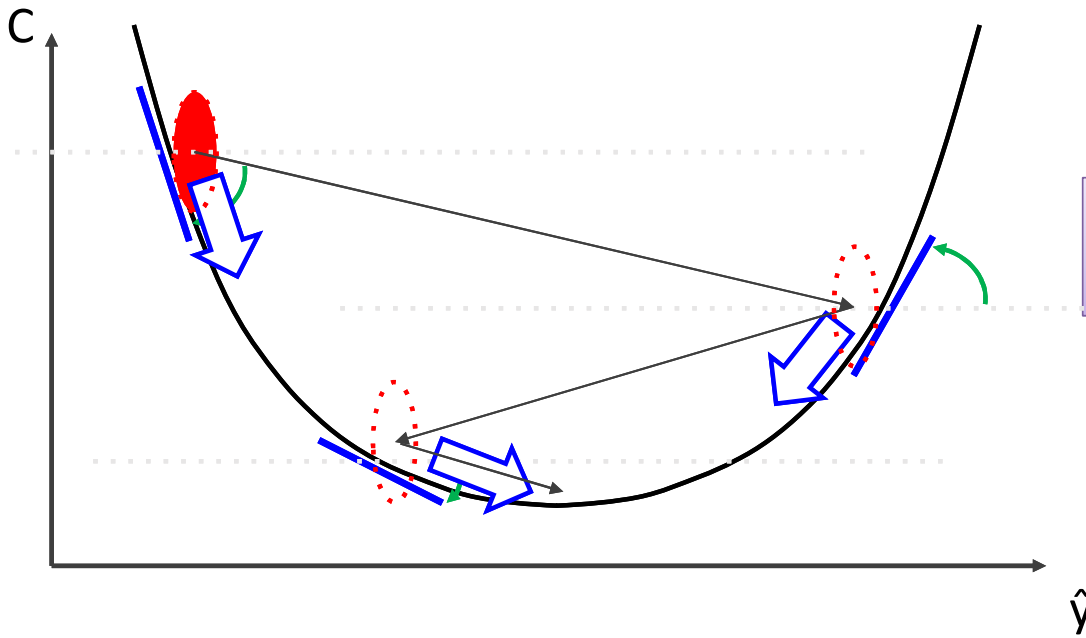
Gradient Descent



Fonte da imagem: neuralnetworksanddeeplearning.com

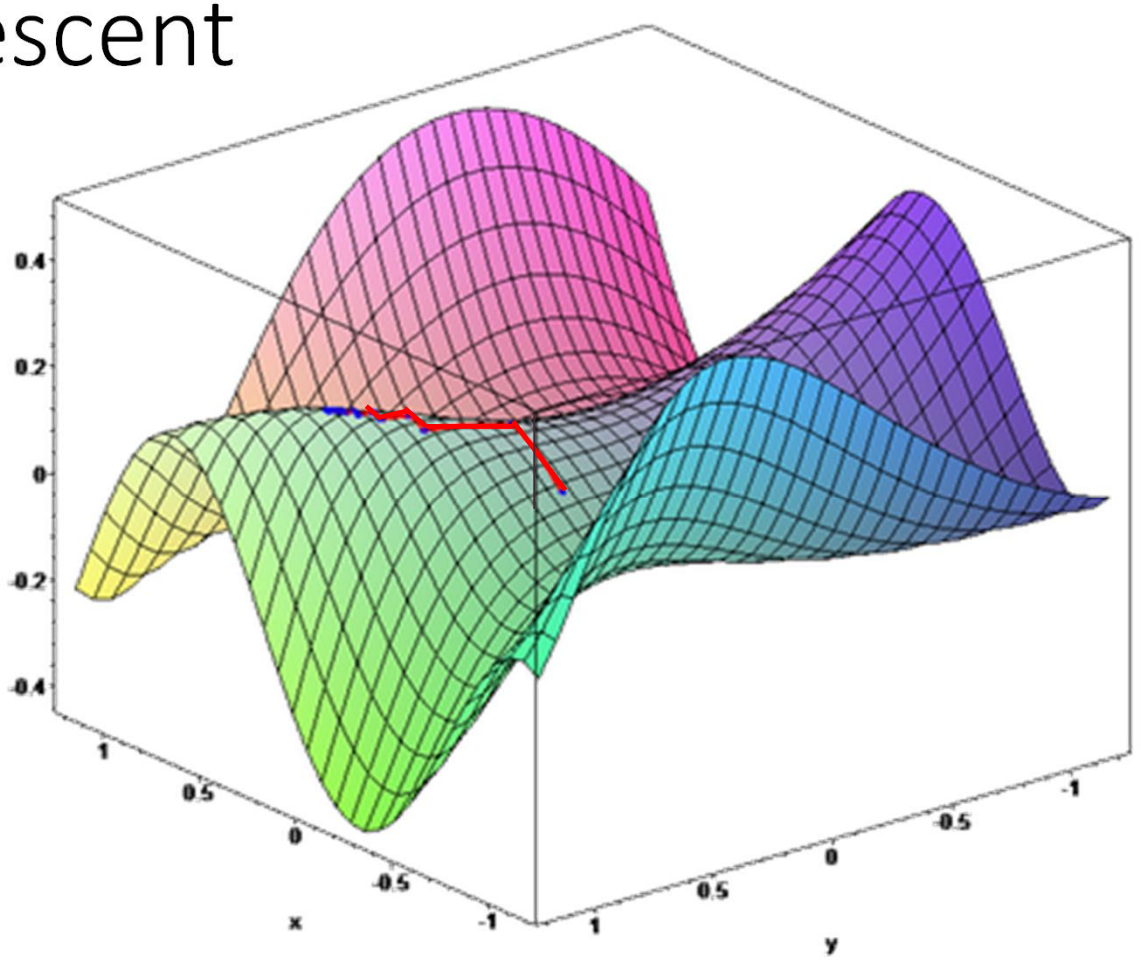
Gradient Descent

Gradient Descent

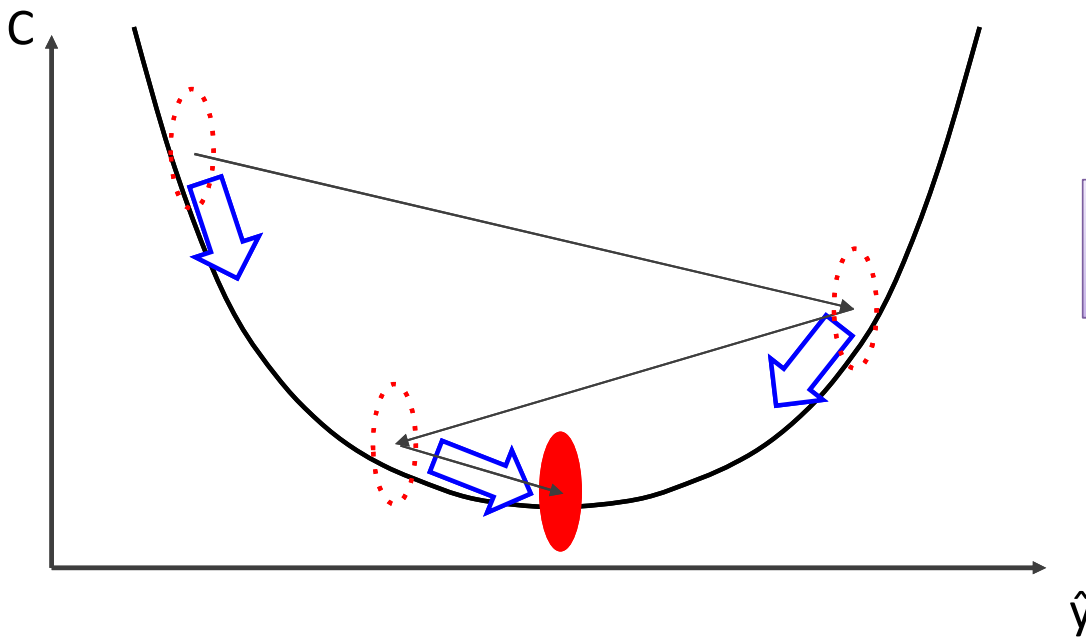


$$C = \frac{1}{2} (\hat{y} - y)^2$$

Gradient Descent



Gradient Descent



$$C = \frac{1}{2} (\hat{y} - y)^2$$

Stochastic Gradient Descent