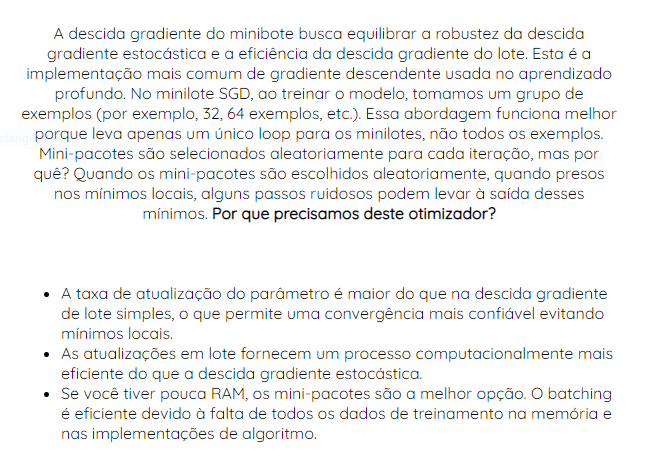
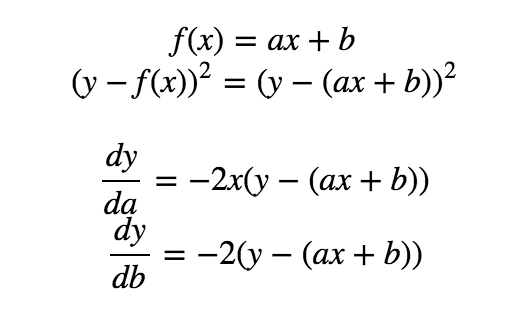
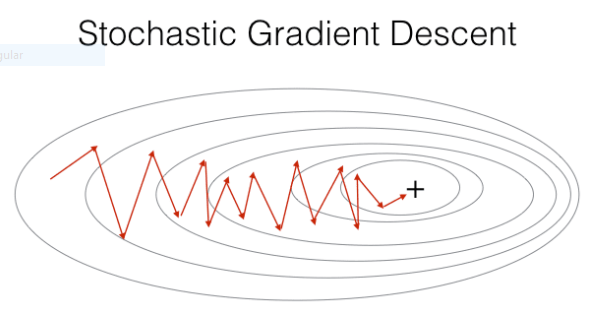
<https://ichi.pro/pt/otimizadores-para-redes-neurais-183947880637191>

<https://medium.com/datarisk-io/tudo-que-voc%C3%AA-j%C3%A1-deveria-saber-sobre-otimiza%C3%A7%C3%A3o-de-hiperpar%C3%A2metros-em-redes-neurais-parte-i-f1d8975f0177>

1. Quais os principais otimizadores usados no treino de redes neurais?
2. Para cada um deles, escreva sua expressão e explique.

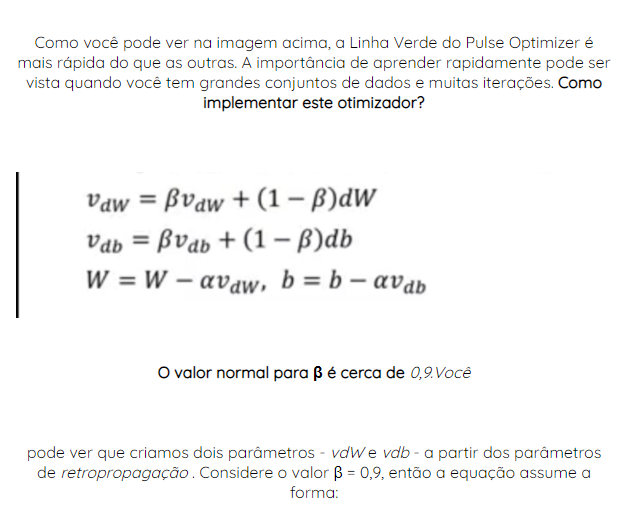
1- SGD:

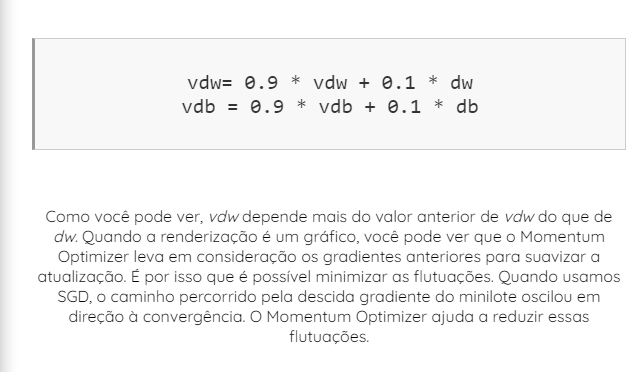


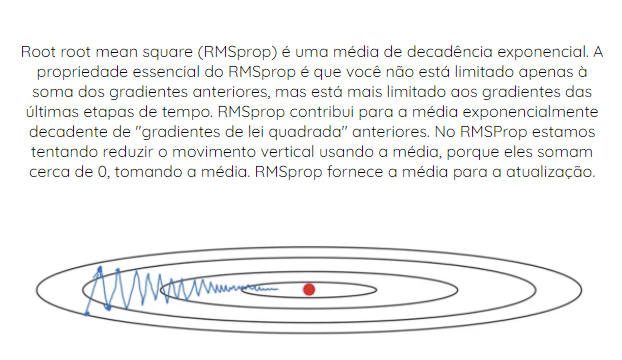


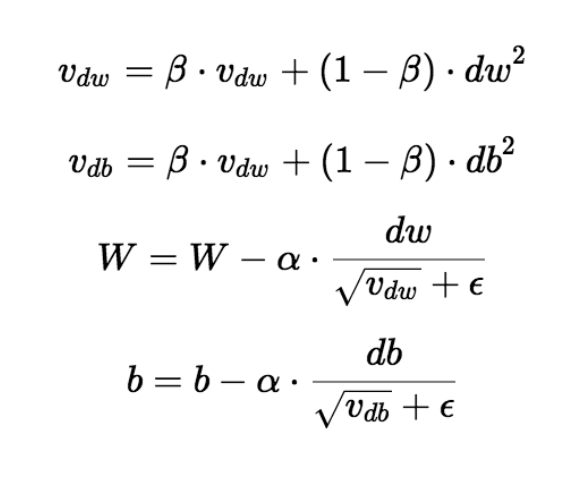
2- Otimizador de impulso:



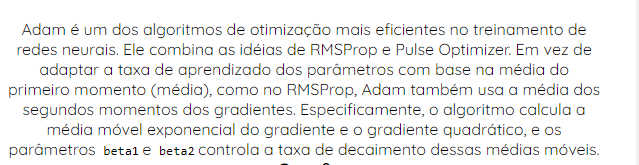


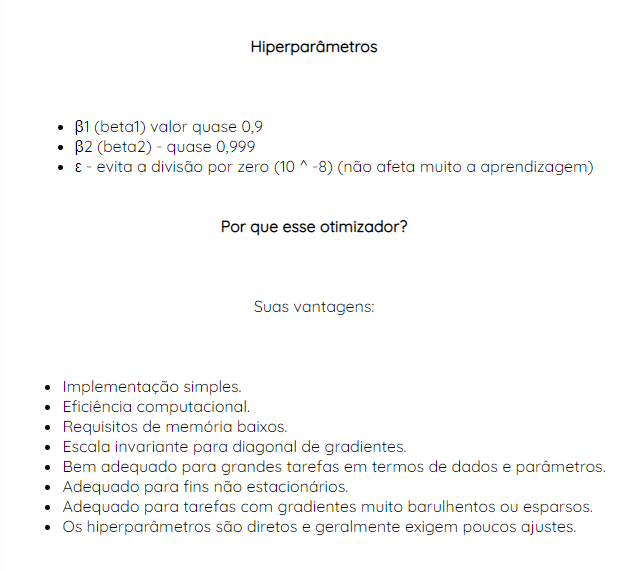


3- RMSprop: 



4- Adam:





1. Quais os prós e contras de cada um deles?

1- SGD: Facilmente interpretável,

Menos complexo, pois é uma derivada de primeira ordem,

Com um grande conjunto de dados, pode levar muito tempo para convergir,

Grande necessidade de memória para calcular o gradiente em todo o conjunto de dados.

2- Otimizador de impulso:

3- RMSprop:

4- Adam:

1. Qual o mais usado e porque?

2) Porque a função de ativação **relu** é a mais usada?

A principal vantagem de usar a **função ReLU** sobre outras funções de **ativação** é que ela não ativa todos os neurônios ao mesmo tempo.

3) No código do fashion - minist, faça as modificações necessárias para que seja possível alterar os valores do learning rate e realize testes com diferentes valores. Use logs para gerar um gráfico comparando o loss para os diferentes valores. Faça o mesmo para o código da regressão.

4) Altere o otimizador SGD para ADAM e repita a questão 3. Os resultados foram melhores?

5) Usar o exemplo do slide e fazer um lista em loop para valores do learning rate