

## Para saber mais

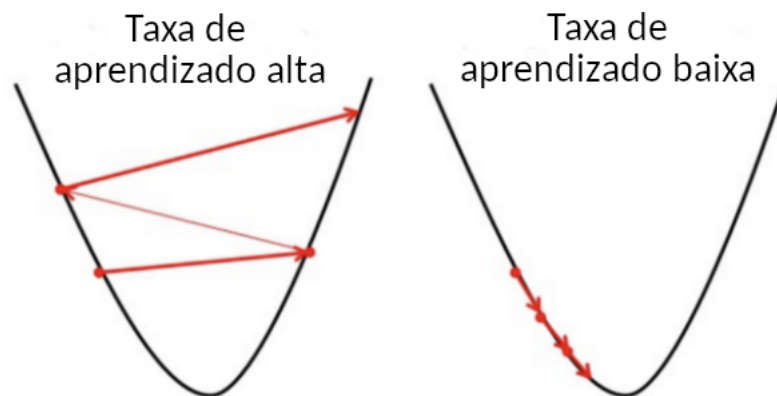
Quando vamos escolher uma taxa de aprendizado para treinar um modelo regressor, devemos escolher um bom valor. Para isto, podemos fazer testes e acompanhar o comportamento da função de custo.

A função de custo no caso, o erro quadrático médio, como é uma função que soma vários valores elevados ao quadrado, seu valor mínimo é zero e o valor máximo pode ser muito alto. Desta forma, o gradiente desta função apontará para o sentido de maior variação.

No método do gradiente descendente, atualizamos o valor de nossos parâmetros, subtraindo o valor do gradiente, fazendo com que nossa função de custo **sempre** diminua de valor a cada iteração, caso a taxa de aprendizado seja pequena o suficiente.

Desta forma, existe uma tendência para escolhermos um valor baixo para a taxa de aprendizado. No entanto, quando escolhemos uma taxa de aprendizado excessivamente baixa, possivelmente menor que  $10^6$ , torna-se necessário executar o gradiente descendente por uma quantidade muito elevada de épocas, tornando o treinamento lento e computacionalmente custoso.

Logo, para acelerar o treinamento, podemos escolher uma taxa de aprendizado maior, sem perder eficiência do algoritmo. Mas devemos nos atentar para o fato de que aumentar o valor da taxa de aprendizado faz com que cada iteração do gradiente descendente aumente o salto na função de custo. Caso a taxa de aprendizado seja grande demais, possivelmente maior que um, a atualização dos parâmetros pode fazer a função de custo divergir.



Na figura acima, vemos na imagem da esquerda o comportamento da função de custo quando a taxa de aprendizado é alta. Você pode notar que o sentido de cada mudança nos parâmetros é correto. No entanto, o tamanho do salto é maior do que o adequado, fazendo a função de custo divergir.

Na imagem da direita, já vemos uma otimização usando uma taxa de aprendizado mais baixa. Desta forma, os parâmetros são atualizados no sentido correto para a minimização e cada iteração faz o valor da função de custo ser menor.

Então, como escolhemos um valor para a taxa de aprendizado? A princípio, devemos experimentar. Colocamos um valor qualquer, por exemplo  $10^{-4}$ , e acompanhamos o valor da função de custo. Caso a função esteja caminhando para o mínimo, podemos tentar aumentá-la. E caso a função de custo esteja divergindo, devemos diminuir a taxa de aprendizado.

