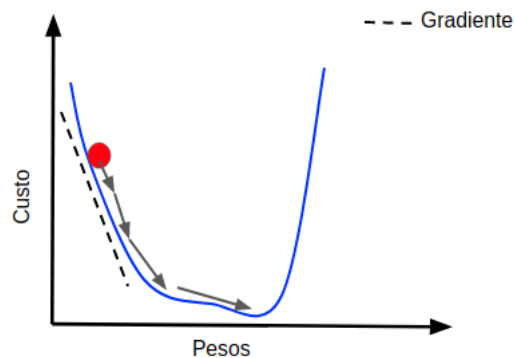


Para saber mais

O algoritmo que usamos para atualizar os pesos da rede recebe o nome de gradiente descendente e é o mais usado em machine learning, embora existam algumas variações dele.

A ideia de aplicação do gradiente descendente é a minimização da função de custo. Isso significa encontrar o ponto onde a função de custo possui o menor valor. Esse valor será mínimo quando o nosso número de bicicletas alugadas estimado for o mais próximo possível do número conhecido de número de bicicletas alugadas. Através da matemática, sabemos que o ângulo de descida de uma função é a derivada da função com respeito a um determinado valor. No nosso caso queremos derivar a função de custo em relação aos pesos, ou em relação ao valor estimado que é dependente dos pesos. Queremos saber em que direção devemos atualizar os pesos para que o custo diminua.



Devemos ter em mente que conforme andamos pela função o ângulo de descida muda. Por isso, o algoritmo de gradiente descendente é aplicado iterativamente. Andamos um passo na função e atualizamos os pesos. Depois disso, andamos mais um passo e verificamos se progredimos. O quanto andamos em cada passo é a taxa de aprendizagem. O tamanho da taxa de aprendizagem é importante. Se escolhermos uma taxa de aprendizagem muito alta, podemos passar direto do ponto de mínimo, sem nunca chegar neste valor. Embora a sua escolha seja muitas vezes por tentativa e erro, a taxa de aprendizagem pode ser escolhida de forma sistemática. Um valor ótimo pode ser obtido como aquele que fornece a maior diferença entre dois valores de custo consecutivos. Além disso, em alguns casos, a taxa de aprendizagem pode ser atualizada no decorrer da aplicação do algoritmo de gradiente descendente. Diminuímos o valor da taxa de aprendizagem aos poucos conforme nos aproximamos do ponto de mínimo da função de custo.