

Teoria T01

Preencha o círculo dos itens verdadeiros e deixe em branco o dos falsos.

- Para que haja garantia de convergência do Perceptron, os dados do problema devem ser linearmente separáveis e a taxa de aprendizado deve ser pequena.
- A importância da Sum of Squared Errors (SSE) como função de custo na Adaline é que por se tratar de uma função não convexa, o algoritmo gradiente descendente (GD) consegue convergir adequadamente.
- Para conjunto de treinamento com uma quantidade enorme de exemplos, o algoritmo GD estocástico leva vantagem sobre o Batch GD, por realizar atualização incremental de pesos.
- Reescala de características com $\mathbf{x}'_j = \frac{\mathbf{x} - \mu_j}{\sigma_j}$ ajuda a acelerar o processo de convergência do GD.
- Podemos utilizar regressão logística para treinar um modelo que estima a nota que você daria a um novo filme lançado nos cinemas.
- Um classificador treinado com menos dados é menos suscetível à superespecialização.
- Um modelo com alta variância é muito sensível aos dados de treinamento e, portanto, se torna superespecializado.
- Um modelo com alto viés é afetado pela mudança de dados de treinamento.
- O uso da regularização L2 permite maior controle sobre overfitting e underfitting de modelos.
- Utilizando abordagem One-vs-All, precisamos realizar n treinamentos para n classes existentes, com $n \geq 3$.