

**Data Science
Academy**

www.datascienceacademy.com.br

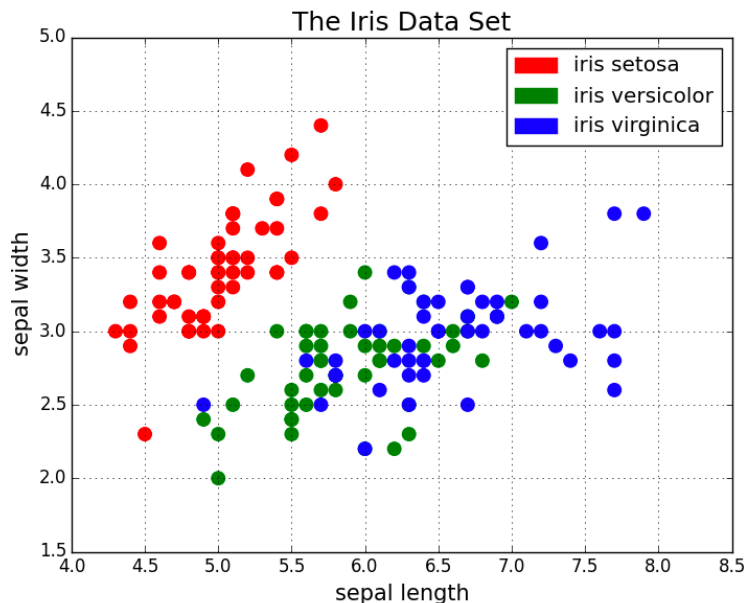
Deep Learning I

Função Softmax

A função de ativação final que examinaremos é a função de ativação softmax, que merece um destaque especial. Junto com a função de ativação linear, a softmax geralmente é encontrado na camada de saída de uma rede neural. A função softmax é usada em uma rede neural de classificação.

O neurônio que possui o valor mais alto reivindica a entrada como membro de sua classe. A função de ativação softmax força a saída da rede neural a representar a probabilidade da entrada se inserir em cada uma das classes. Sem a softmax, as saídas do neurônio são simplesmente valores numéricos, com o mais alto indicando a classe vencedora.

Para compreender como a função de ativação softmax é usada, analisaremos um problema comum de classificação da rede neural. O conjunto de dados íris contém quatro medidas para 150 flores diferentes. Cada uma dessas flores pertence a uma das três espécies de íris. Quando você fornece as medidas de uma flor, a função softmax permite que a rede neural lhe dê a probabilidade de que essas medidas pertençam a cada uma das três espécies. Por exemplo, a rede neural pode dizer-lhe que há 80% de chance de que a íris seja *setosa*, uma probabilidade de 15% de ser *virginica* e apenas uma probabilidade de 5% para *versicolor*. Por se tratar de probabilidades, elas devem somar 100%. Não poderia haver uma probabilidade de 80% de *setosa*, uma probabilidade de 75% de *virginica* e uma probabilidade de 20% de *versicolor* - esse tipo de resultado seria sem sentido.



Para classificar dados de entrada em uma das três espécies de íris, você precisará de um neurônio de saída para cada uma das três espécies. Os neurônios de saída não especificam inerentemente a probabilidade de cada uma das três espécies. Portanto, é desejável fornecer probabilidades que somam 100%. A rede neural irá dizer-lhe a probabilidade de uma flor ser cada uma das três espécies. Para obter a probabilidade, usamos a função softmax:

$$\phi_i = \frac{e^{z_i}}{\sum_{j \in \text{group}} e^{z_j}}$$

Na equação acima, i representa o índice do neurônio de saída (o) sendo calculado, e j representa os índices de todos os neurônios no grupo / nível. A variável z designa a matriz de neurônios de saída. É importante notar que a ativação softmax é calculada de maneira diferente das outras funções de ativação. Quando a softmax é a função de ativação, a saída de um único neurônio é dependente dos outros neurônios de saída. Podemos observar na equação, que a saída dos outros neurônios de saída está contida na variável z , já que nenhuma das outras funções de ativação neste capítulo utiliza z .

Considere uma rede neural treinada que classifique os dados em três categorias, como as três espécies de íris. Nesse caso, você usaria um neurônio de saída para cada uma das classes alvo. Considere que a rede neural deve produzir o seguinte:

Neurônio 1: setosa: 0.9

Neurônio 2: virginica: 0.4

Neurônio 3: versicolor: 0.2

A partir da saída acima, podemos ver claramente que a rede neural considera que os dados representam uma íris setosa. No entanto, esses números não são probabilidades. O valor 0.9 não representa uma probabilidade de 90% dos dados que representam um setosa. Estes valores somam 1,5. Para que eles sejam tratados como probabilidades, eles devem somar 1.0. O vetor de saída para esta rede neural é o seguinte:

[0.9, 0.4, 0.2]

Se este vetor for fornecido à função softmax, o seguinte vetor é retornado:

[0.47548495534876745, 0.28839620365112, 0.2361188410001125]

Os valores acima somam 1.0 e podem ser tratados como probabilidades. A probabilidade de os dados representarem uma íris setosa é de 48% porque o primeiro valor no vetor gira em torno de 0,48 (48%).

Por esta razão, usamos a função de ativação softmax na camada de saída de redes neurais de classificação.