



**Data Science
Academy**

www.datascienceacademy.com.br

Deep Learning I

Cross-Entropy



Para treinar uma rede neural, você precisa encontrar o erro entre as saídas calculadas e as saídas alvo desejadas. A medida de erro mais comum é chamada de erro quadrático médio (ou Mean Squared Error). No entanto, existem alguns resultados de pesquisa que sugerem o uso de uma medida diferente, denominada erro de entropia cruzada, como método preferível em relação ao erro quadrático médio.

A medida de entropia cruzada tem sido utilizada como alternativa ao erro quadrático médio. A entropia cruzada pode ser usada como uma medida de erro quando as saídas de uma rede podem ser pensadas como representando hipóteses independentes (por exemplo, cada nó significa um conceito diferente) e as ativações dos nós podem ser entendidas como representando a probabilidade (ou a confiança) que cada uma das hipóteses pode ser verdadeira. Nesse caso, o vetor de saída representa uma distribuição de probabilidade, e nossa medida de erro - entropia cruzada - indica a distância entre o que a rede acredita que essa distribuição deve ser e o que realmente deveria ser. Existe também uma razão prática para usar a entropia cruzada. Pode ser mais útil em problemas nos quais os alvos são 0 e 1. A entropia cruzada tende a permitir que erros alterem pesos mesmo quando houver nós saturados (o que significa que suas derivadas são próximas de 0). Vejamos a entropia cruzada em mais detalhes:

A entropia cruzada é comumente usada para quantificar a diferença entre duas distribuições de probabilidade. Geralmente, a distribuição "verdadeira" (dos dados usados para treinamento) é expressa em termos de uma distribuição One-Hot.

Por exemplo, suponha que para uma instância de treinamento específica (uma única linha no seu dataset), a classe seja B (de 3 possíveis possibilidades: A, B e C). A distribuição única para esta instância de treinamento é, portanto:

Pr(Class A)	Pr(Class B)	Pr(Class C)
0.0	1.0	0.0

Você pode interpretar a distribuição acima da seguinte forma: a instância de treinamento tem 0% de probabilidade de ser classe A, 100% de probabilidade de ser classe B e 0% de probabilidade de ser a classe C.

Agora, suponha que seu algoritmo de aprendizado de máquina tenha previsto a seguinte distribuição de probabilidade:

Pr(Class A)	Pr(Class B)	Pr(Class C)
0.228	0.619	0.153

Quão próxima é a distribuição prevista da distribuição verdadeira? É isso que determina o erro de entropia cruzada. A entropia cruzada é representada por esta fórmula:



$$H(p, q) = - \sum_x p(x) \log q(x).$$

A soma é sobre as três classes A, B e C. Se você completar o cálculo, você achará que a perda é 0.479. Então, é assim que "longe" está a sua previsão da distribuição verdadeira.

A entropia cruzada é uma das muitas funções de perda possíveis. Essas funções de perda são tipicamente escritas como $J(\theta)$ e podem ser usadas dentro da descida do gradiente, que é uma estrutura iterativa para mover os parâmetros (ou coeficientes) para os valores ótimos. A entropia cruzada descreve a perda entre duas distribuições de probabilidade.

Ao usar uma rede neural para realizar classificação e predição, geralmente é melhor usar o erro de entropia cruzada do que o erro de classificação e um pouco melhor usar o erro de entropia cruzada do que o erro quadrático médio para avaliar a qualidade da rede neural. É importante deixar claro que estamos lidando apenas com uma rede neural que é usada para classificar os dados, como a previsão da afiliação de um partido político (democrata, republicana, outro) de dados independentes, como idade, sexo, renda anual, e assim por diante. Não estamos lidando com uma rede neural que faz regressão, onde o valor a ser previsto é numérico.