

1. Soma dos Erros Quadráticos:

$$\sum_{i=1}^3 (y_i - \hat{y}_i)^2 = 1 + 4 + 9 = 14$$
2. Divida pelo número de amostras ($N = 3$):

$$MSE = \frac{1}{3} \times 14 = 4.67$$

Portanto, o MSE para essas previsões é 4.67.

Aqui está uma imagem que ilustra o conceito do MSE:

User
5. Função de Perda - Entropia Cruzada (Cross-Entropy)

Model
Excelente! A Entropia Cruzada (Cross-Entropy) é outra função de custo extremamente importante, mas ao contrário do MSE, ela é usada principalmente em problemas de **classificação**, onde o objetivo é prever a probabilidade de uma amostra pertencer a uma ou mais classes.

Propósito da Entropia Cruzada:

A Entropia Cruzada mede o desempenho de um modelo de classificação cuja saída é uma probabilidade (um valor entre 0 e 1). Ela quantifica a diferença entre a distribuição de probabilidade prevista pelo modelo e a distribuição de probabilidade verdadeira (os rótulos reais).

Quanto maior a Entropia Cruzada, maior a diferença entre as distribuições e pior o desempenho do modelo. O objetivo do treinamento é **minimizar a Entropia Cruzada**.

Tipos de Entropia Cruzada:

Existem duas formas principais, dependendo do problema de classificação:

1. **Entropia Cruzada Binária (Binary Cross-Entropy - BCE)**: Usada para problemas de **classificação binária**, onde há apenas duas classes (por exemplo, "sim/não", "positivo/negativo", "doente/saudável").
2. **Entropia Cruzada Categórica (Categorical Cross-Entropy - CCE)** ou **Entropia Cruzada Multi-Classe**: Usada para problemas de **classificação multi-classe**, onde há três ou mais classes (por exemplo, "gato", "cachorro", "pássaro").

1. Entropia Cruzada Binária (BCE)

A Fórmula da BCE:

Para uma única amostra, onde y é o rótulo verdadeiro (0 ou 1) e \hat{y} é a probabilidade prevista pelo modelo para a classe 1 (a probabilidade de ser 1):

$$BCE = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

Para um conjunto de N amostras, o custo médio é:

$$BCE = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

Onde:

- y_i é o rótulo verdadeiro da i -ésima amostra (0 ou 1).
- \hat{y}_i é a probabilidade prevista para a classe 1 pela i -ésima amostra.
- \log é o logaritmo natural.

Como funciona:

- Se $y = 1$: A fórmula se simplifica para $-\log(\hat{y})$. Queremos que \hat{y} seja próximo de 1 (para que $-\log(\hat{y})$ seja próximo de 0). Se \hat{y} for próximo de 0, $-\log(\hat{y})$ será muito grande.
- Se $y = 0$: A fórmula se simplifica para $-(1 - \hat{y})$. Queremos que \hat{y} seja próximo de 0 (para que $1 - \hat{y}$ seja próximo de 1, e $-(1 - \hat{y})$ seja próximo de 0). Se \hat{y} for próximo de 1, $1 - \hat{y}$ será próximo de 0, e $-(1 - \hat{y})$ será muito grande.

Exemplo de Cálculo (BCE para uma amostra):

Suponha que o rótulo verdadeiro seja $y = 1$, e o modelo previu $\hat{y} = 0.8$.
 $BCE = -[1 \times \log(0.8) + (1 - 1) \times \log(1 - 0.8)]$
 $BCE = -[\log(0.8)] = 0 \times \log(0.2)$
 $BCE = -\log(0.8) = -(-0.223) \approx 0.223$

Suponha que o rótulo verdadeiro seja $y = 1$, e o modelo previu $\hat{y} = 0.2$ (uma previsão ruim).
 $BCE = -[1 \times \log(0.2) + (1 - 1) \times \log(1 - 0.2)]$
 $BCE = -\log(0.2) \approx -(-1.39) \approx 1.39$

Observe como a BCE é muito maior quando a previsão é ruim.

2. Entropia Cruzada Categórica (CCE)

A Fórmula da CCE:

Para uma única amostra em um problema com K classes:

$$CCE = -\sum_{k=1}^K y_{ik} \log(\hat{y}_{ik})$$

Para um conjunto de N amostras, o custo médio é:

$$CCE = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_{ik} \log(\hat{y}_{ik})$$

Onde:

- y_{ik} é 1 se a i -ésima amostra pertence à k -ésima classe, e 0 caso contrário (codificação one-hot).
- \hat{y}_{ik} é a probabilidade prevista pelo modelo para a i -ésima amostra pertencer à k -ésima classe (geralmente a saída de uma camada Softmax).

Como funciona:

Devido à codificação one-hot de y_{ik} , apenas o termo onde $y_{ik} = 1$ (a classe verdadeira) permanecerá na soma. Todos os outros termos serão zero. Portanto, a CCE penaliza o modelo pela baixa probabilidade atribuída à classe verdadeira.

Exemplo de Cálculo (CCE para uma amostra):

Suponha 3 classes. O rótulo verdadeiro é a classe 1 (cachorro), representado em one-hot como $y = [1, 0, 0]$. O modelo previu as probabilidades $\hat{y} = [0.7, 0.2, 0.1]$ (usando Softmax).

$$CCE = -[y_1 \log(\hat{y}_1) + y_2 \log(\hat{y}_2) + y_3 \log(\hat{y}_3)]$$

$$CCE = -[1 \times \log(0.7) + 0 \times \log(0.2) + 0 \times \log(0.1)]$$

$$CCE = -\log(0.7) \approx -(-0.357) \approx 0.357$$

Agora, se o modelo tivesse previsto mal, por exemplo, $\hat{y} = [0.1, 0.8, 0.1]$ (atribuindo alta probabilidade à classe errada):

$$CCE = -[1 \times \log(0.1) + 0 \times \log(0.8) + 0 \times \log(0.1)]$$

$$CCE = -\log(0.1) \approx -(-2.203) \approx 2.203$$

Novamente, o custo é muito maior quando a previsão para a classe verdadeira é baixa.

Aqui está uma imagem que ilustra o conceito geral da Entropia Cruzada, mostrando como ela mede a "distância" entre a distribuição real e a prevista:



Start typing a prompt

Run Ctrl+←