Loss functions

A loss function is a mathematical function that measures how wrong the model's prediction are. It compares the prediction with actual value.

$\mathcal{L}(y, \hat{y})$; the model tries to :

$$\min_{\theta} \mathcal{L}(y, f(x;\theta)).$$

Loss function in Regression

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Penalizes large error heavily because of the square term. Used in stable problems where outliers are not common.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

It penalizes error linearly. Less sensitive to outliers than MSE and is good for noisy data where outliers should not dominate.

Huber loss :- Combination of both.

$$L = \begin{cases} \frac{1}{2}(y-\hat{y})^2 & \text{if } |y-\hat{y}| < \delta \\ \delta|y-\hat{y}| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}$$

Behaves like MSE for small errors
MAE for large errors.

Loss functions for classification

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

$\rightarrow$ Binary class.

for multi-class cases,

$$p_k = \frac{e^{z_k}}{\sum_j e^{z_j}}$$

loss $$CE = -\sum_{k=1}^{K} y_k \log(p_k).$$

—— x ——

Categorical Cross Entropy Loss

Multi class classification problems. It measures the performance of a classification model whose output is probability distribution over multiple classes.

$$CCE = -\sum_{i=1}^{n}\sum_{j=1}^{K} y_{ij} \log(\hat{y}_{ij}).$$

* labels are one-hot encoded

$K \rightarrow$ no of classes

Sparse Categorical Cross-Entropy loss

Multiclass classification

Labels are integers.

$$-\sum_{i=1}^{n} \log(\hat{y}_{i,y_i})$$

y_i is the integer representing the correct class for datapoint i

Example

true class = class 2 (out of 3)

$y = [0, 1, 0]$

$\hat{y} = [0.1, 0.7, 0.2]$

$L = -\log(0.7)$

true class = 1

$\hat{y} = [0.1, 0.7, 0.2]$

$L = -\log(0.7)$

Categorical Cross entropy uses one-hot encoded labels, whereas sparse categorical cross entropy uses integer Label. Both compute same loss but differ in label representation.

Hinge Loss :- Alternative loss function for binary classification effective for SVMs as it used for maximum-margin classifier.

$$\boxed{\text{Hinge loss} = \max(0, 1 - y \cdot f(x))}$$

true labels for single training example (x, y)

$y \in \{-1, +1\}$; $f(x) \rightarrow$ model's predicted score.

if the prediction is correct and confident i.e $f(x) \geq 1$ $\rightarrow$ loss $= 0$

if the prediction is correct but not confident enough $\rightarrow$ some loss

if the prediction is wrong $\rightarrow$ large loss.

Example : $y = +1$

$f(x) = 0.6$; loss $= \max(0, 1 - 1 \times 0.6)$
$= 0.4$

But if $f(x) = 1.3$ loss $= 0$.

Note: $f(x) = W^T x + b$ $\quad f(x) \in (-\infty, +\infty)$

for positive $\rightarrow$ predicts class $+1$

negative $\rightarrow$ predicts class -1

magnitude $\rightarrow$ confidence.

Triplet Loss :- Triplet loss is used to learn embeddings by comparing the relative distances between triplets: anchor, positive example, negative example.

$$\frac{1}{N} \sum_{i=1}^{N} \left[\| f(x_i^a) - f(x_i^p) \|_2^2 - \| f(x_i^a) - f(x_i^n) \|_2^2 + \alpha \right]_+$$