Loss for Classification and Regression

ML Instruction Team, Fall 2022

CE Department
Sharif University of Technology

Classification Loss

- Both the cross-entropy and the Kullback–Leibler (KL) divergence measures the distance between two probability distributions P and Q.

$$H(P, Q) = -\sum_x P(x) \log Q(x)$$

$$KL(P \mid Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} = H(P, Q) - H(P)$$

- In the above formula, $H(P) = H(P, P)$ is the entropy of the distribution P, which is a constant term.
- Hence, it turns out that the minimization of KL divergence is equivalent to the minimization of cross-entropy.

Classification Loss

- Logarithmic loss indicates how close a distribution of prediction probability comes to the real probability distribution of the data in binary classification.

$$H(p,q) = -\frac{1}{m}\sum_{i=1}^{m} y_i \log(p(y_i)) + (1-y_i)\log(1-p(y_i))$$

where $y_i \in \{0,1\}$

- Cross-entropy loss indicates how close a distribution of prediction probability comes to the real probability distribution of the data in multilabel classification.

$$H(p,q) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k} y_j^{(i)} \log(p(y_j^{(i)}))$$

where $y_j^{(i)}$ and $p(y_j^{(i)})$ is respectively the true and predicted probability of the class i-th of the sample j-th.

Classification Loss

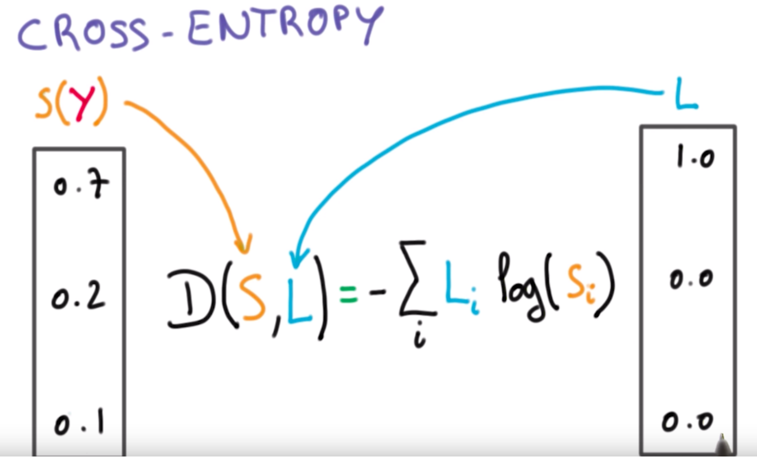

Figure: Cross Entropy Evaluation, Source

Regression Loss

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

$$\text{MAE}(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} |(y_i - \hat{y}_i)|$$

$$\text{MAPE}(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} |\frac{(y_i - \hat{y}_i)}{y_i}|$$

$$\text{Logcosh}(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} \cosh(y_i - \hat{y}_i)$$

$$L_\delta(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{for } |y - \hat{y}| \le \delta \\ \delta \cdot (|(y - \hat{y})| - \frac{1}{2}\delta), & \text{otherwise} \end{cases}$$

Regression Loss

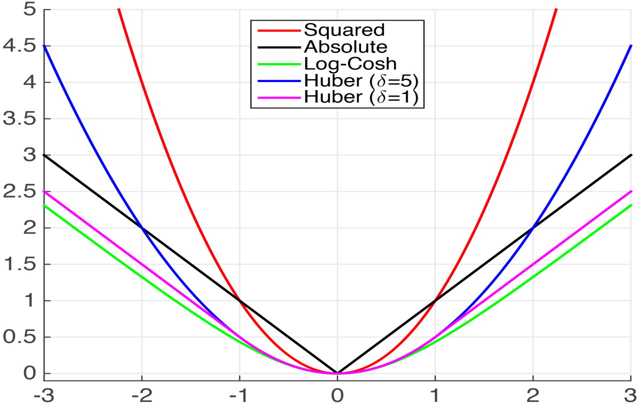

Figure: Various Regression Losses, Source

Thank You!

Any Question?