Cross-Entropy ($ce(y, \hat{y})$)

$$ce(y, \hat{y}) = -\sum_{i=0}^{n} y_i \log(\hat{y}_i)$$

- n : number of classes
- y : actual output
- $\hat{y}$: predicted output

$$\frac{\partial ce(y, \hat{y})}{\partial y_k} = \frac{\partial}{\partial y_k} \sum_{i=0}^{n} y_i \log(\hat{y}_i) =$$

$$= \frac{\partial}{\partial y_k} y_k \log(\hat{y}_k) = y_k \cdot \frac{1}{\hat{y}_k}$$

$$\frac{\partial ce(y, \hat{y})}{\partial y_k} = \frac{y_k}{\hat{y}_k}$$

MSE (mse(y, ŷ))

$$mse(y, \hat{y}) = -\sum_{i=0}^{n} (\hat{y} - y)^2$$

- n : number of classes
- y : actual output
- $\hat{y}$: predicted output

$$\frac{\partial mse(y, \hat{y})}{\partial y_k} = \frac{\partial}{\partial y_k} \sum_{i=0}^{n} (\hat{y}_k - y_k)^2 =$$

$$= \frac{\partial (\hat{y}_k - y_k)^2}{\partial y_k} =$$

$$= 2 \cdot (\hat{y}_k - y_k) \cdot \frac{\partial}{\partial y_k} (\hat{y}_k - y_k) =$$

$$= 2 \cdot (\hat{y}_K - y_K) \cdot (-1)$$

$$\boxed{\frac{\partial mse(y, \hat{y})}{\partial y_K} = 2(y_K - \hat{y}_K)}$$