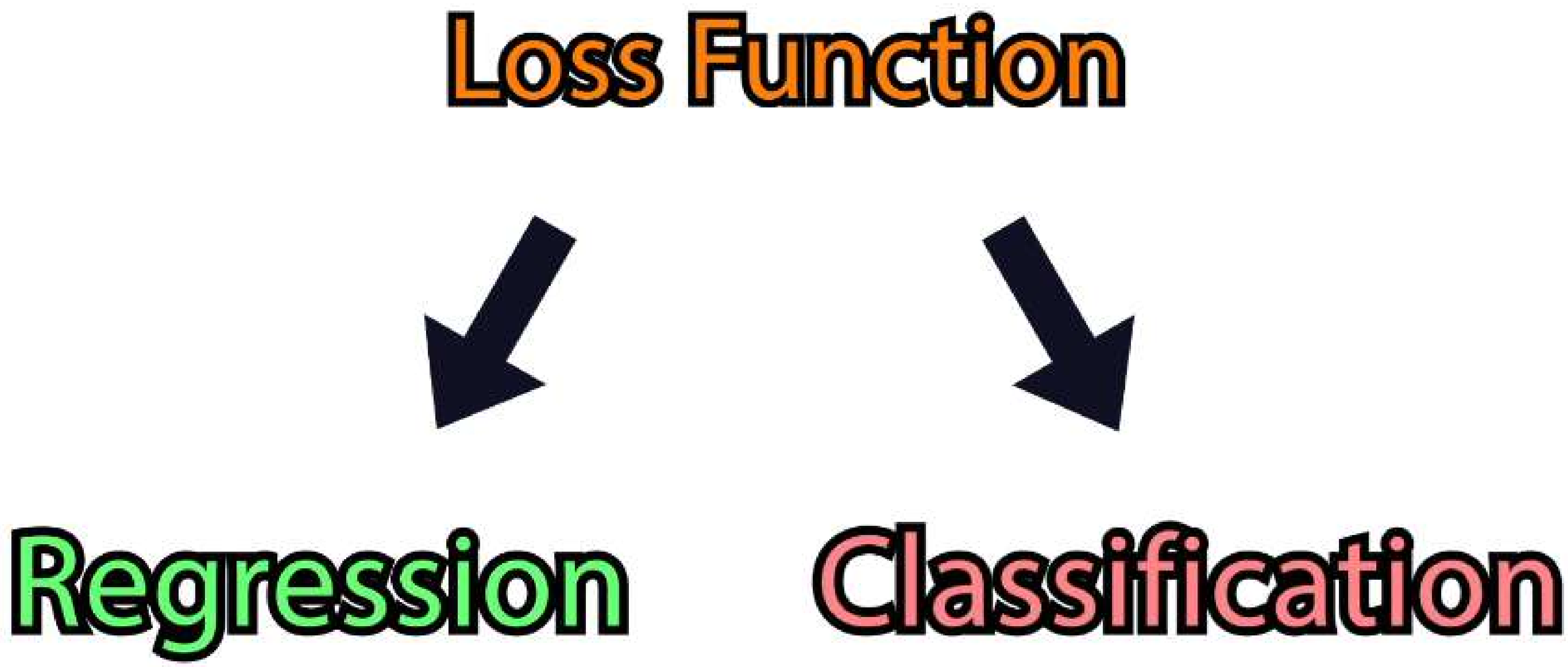
Loss Function

$$L(y,\hat{y}) = \text{Compare}(y,\hat{y})$$

y - Target

$\hat{y}$ - Prediction


Loss Function
Regression
Classification

Regression

y	$\hat{y}$	$y - \hat{y}$	$\lvert y - \hat{y} \rvert$	$(y - \hat{y})^2$
120	154	-34	34	1156
100	113	-13	13	169
110	107	3	3	9
			16.(6)	444.(6)

Regression

y	$\hat{y}$	$y - \hat{y}$	MAE	MSE
120	154	-34	34	1156
100	113	-13	13	169
110	107	3	3	9
			16.(6)	444.(6)

If: $y_1 - \hat{y}_1 = x$

$y_2 - \hat{y}_2 = 2x$

$MAE(y_1, \hat{y}_1) = x$ $MSE(y_1, \hat{y}_1) = x^2$

$MAE(y_2, \hat{y}_2) = 2x$ $MSE(y_2, \hat{y}_2) = 4x^2$

If:

$$y_1 - \hat{y}_1 = x$$

$$y_2 - \hat{y}_2 = 2x$$

$$MSE(y_2, \hat{y}_2) = (2x)^2 = 4x^2$$

$$MSE(y_1, \hat{y}_1) = (x)^2 = x^2$$

Classification

(Win = 1, Loss = 0)

$$L(y,\hat{y}) = \text{Compare categories}$$

y - True category

$\hat{y}$ - Predicted category

Classification

(Win = 1, Loss = 0)

Binary Cross Entropy

$$L_{(y,\hat{y})} = -\underbrace{\frac{1}{n}\sum_{i=1}^{n}}_{\text{Averaging for all observations}}\left(y_i \cdot \log \hat{y}_i + (1-y_i) \cdot \log(1-\hat{y}_i)\right)$$

Predictions are rounded using some threshold like 0.5

$$L(y_i, \hat{y}_i) = -\, y_i \cdot \log \hat{y}_i - (1-y_i) \cdot \log(1-\hat{y}_i)$$

$$L(1, 0.3) = 1.2$$

$$L(0, 0.8) = 1.61$$

$$L(1, 0.9) = -1 \cdot (-0.1) - 0 \cdot (-2.3)$$

Multi-Class Classification

$$L(y,\hat{y}) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{C} y_{ij}\cdot\log(\hat{y}_{ij})$$

$$L(y_i,\hat{y}_i) = -\sum_{j=0}^{C-1} y_{ij}\cdot\log(\hat{y}_{ij}) =$$

$$= -0\cdot\log(0.04) - 1\cdot\log(0.1) - 0\cdot\log(0.86)$$

$$= 2.3$$

Note: cross entropy is a harder topic. We will refer to it separately!

Ex. $C=3$

Labels $= 0,1,2$

$y_0 : [1,0,0]$
$y_1 : [0,1,0]$
$y_2 : [0,0,1]$

$\hat{y}_0 = 0.04$
$\hat{y}_1 = 0.1$
$\hat{y}_2 = 0.86$

Categorical Cross-Entropy loss.

Sigmoid Binary

Actual.

1 ✓ Belong.

0 - Not Belong.

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

Softmax.

Prediction

$$\frac{\log_e .7 + \log_e .8 + \log_e .9}{}$$

$$\frac{-1(-.66)}{3} = .23$$

Data Points	Class X	Class Y	Class Z	Actual Class
1	.7 ×1	.2 ×0	.3 ×0	X ✓
2 —	.1 ×0	.8 ×1	.2 ×0	Y ✓
3	.2 ×0	.1 ×0	.9 ×1	Z ✓

Binary Cross-Entropy (Log Loss)

$\ln(x)$

$\log_2 x$

1, 0

Actual   pred.

$\frac{1}{x(\ln 2)}$

$$BCE = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log_e(\hat{y}_i) + (1 - y_i)\log_e(1 - \hat{y}_i)\right]$$

$1 \times \log_e(.9) + (1-0)(\log_e(1-.7)$

$\ln .3$

Email	Actual label	Model Prediction
A	1 (SPAM)	.9 → $\ln(.9) +$
B	0 (NOT SPAM)	.7 $\ln(.3) +$
C	1 (SPAM)	.4 $\ln(.4)$