4.4 Loss Function in Predictive Model.

4.4.1. In Regression:

$L(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$: Square loss

$L(\hat{y}, y) = |\hat{y} - y|$: Absolute loss

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

4.4.2 In Classification:

4.4.2.1 Non-differentiable loss function

- Assume k different classes, loss function $L(\hat{y}, y)$ can be represented as $K \times K$ matrix.

$$L = \begin{bmatrix} L(1,1) & L(1,2) & \cdots & L(1,k) \\ L(2,1) & L(2,2) & \cdots & L(2,k) \\ \vdots & \vdots & & \vdots \\ L(k,1) & L(k,2) & \cdots & L(k,k) \end{bmatrix}$$

- the diagonal $L(i,i)$ are zero as correct prediction.
- the off-diagonal $L(i,j)$ are positive : loss incurred when predicting "i" instead of "j"

- Zero-one loss
  - If $L(i,j) = 1$ for $i \neq j$, 0 otherwise : $L(i,j) = \begin{cases} 1 & i \neq j \\ 0 & i = j \end{cases}$
  - The expected $J(h) = E_{x,y}[L(h(x), y)]$

$$= \sum_{i,j} L(i,j) P(i,j) \quad , \text{where } P(i,j) = P(\hat{y}=i, y=j)$$

$$= \sum_{i \neq j} P(i,j)$$ misclassification or error rate.

- Binary Classification

Meaning	$\hat{y}$	y	conditional prop.	other names	meaning	joint probability
TRUE-positive rate	1	1	$P(\hat{y}=1 \mid y=1)$	sensitivity, hit rat re call	true positive	$P(\hat{y}=1, y=1)$
true-negative rate	-1	-1	$P(\hat{y}=-1 \mid y=-1)$	specificity	true negative	$P(\hat{y}=-1, y=-1)$
false-positive rate	1	-1	$P(\hat{y}=1 \mid y=-1)$	type 1 error	false positive	$P(\hat{y}=1, y=-1)$
false-negative rate	-1	1	$P(\hat{y}=-1 \mid y=1)$	type 2 error	false negative	$P(\hat{y}=-1 \mid y=+1)$