Artificial Intelligence (AI)

Lec06: Logistic Regression Part 2

충북대학교

문성태 (지능로봇공학과)

stmoon@cbnu.ac.kr

03

Evaluation

Evaluation: Accuracy

- Classification 문제로 성능은 정확도 (accuracy)로 평가
- 데이터 형태에 따라 accuracy로 측정하기 어려운 경우 발생

Evaluation

- Accuracy
- Precision
- Recall

Evaluation

0	0	
0	0	
0	0	
0	0	
0	0	
0	1	
1	0	
1	1	
1	0	
0	0	

		Actual	
		Positive	Negative
Predicted	Positive	**TP**	**FP**
	Negative	**FN**	**TN**

Confusion Matrix

Evaluation

0	0	TN
0	0	TN
0	0	TN
0	0	TN
0	0	TN
0	1	FP
1	0	FN
1	1	TP
1	0	FN
0	0	TN

❖ Accuracy (정확도)

Accuracy

$$\frac{TP + TN}{TP + FP + FN + TN}$$

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

❖ Precision (정밀도)

Precision

$$\frac{TP}{TP + FP}$$

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

❖ Recall (재현율)

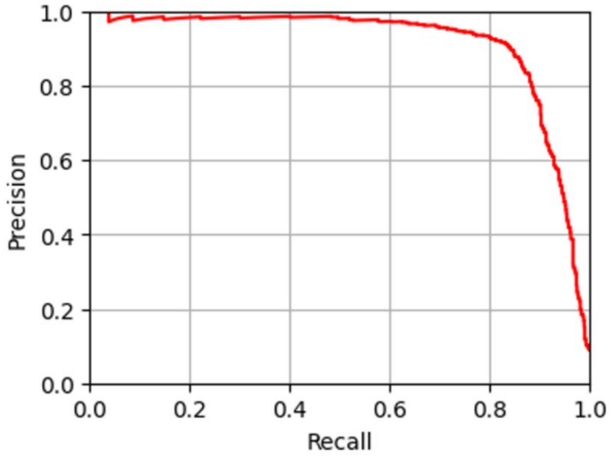
Recall

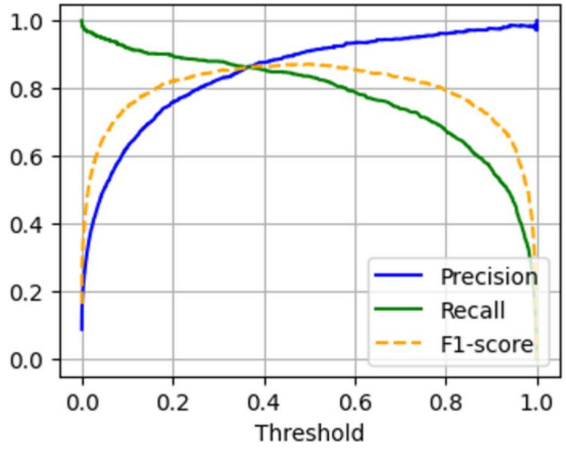
$$\frac{TP}{TP + FN}$$

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

Evaluation

PR curve





❖ F1 (조화 평균 두 값의 비율을 중시 여김)

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

참고) 비교 대상이 서로 역수의 관계가 있을때 조화 평균 활용

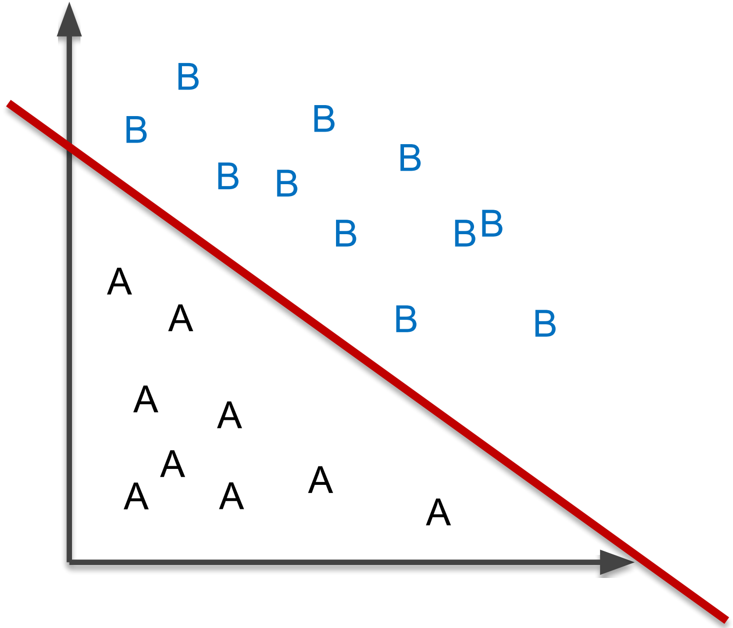
04

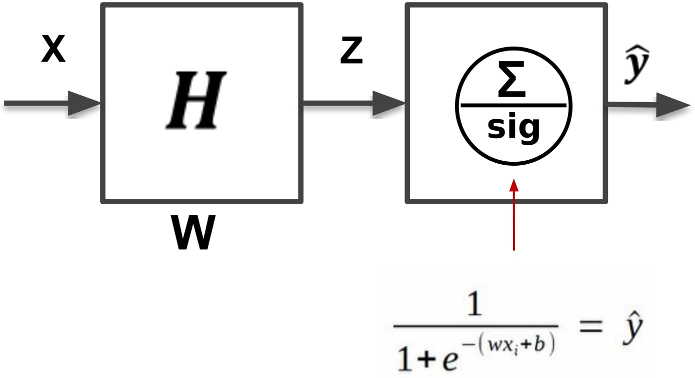
Multinomial Logistic Regression

(Softmax Regression)

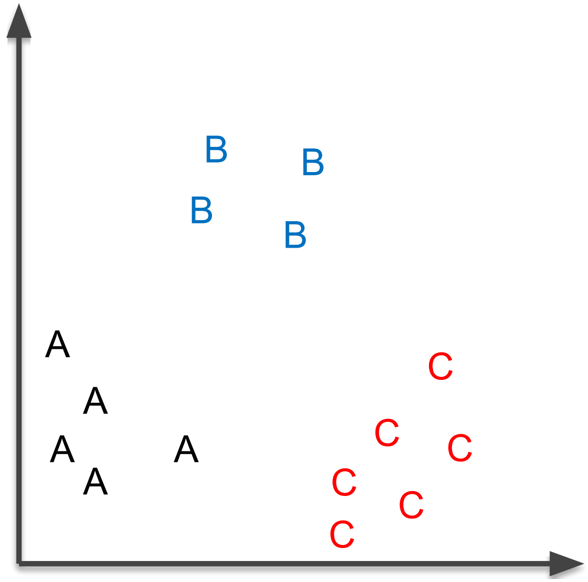
Recap: Logistic Regression for binary classification

- 이진 분류된 영역에서 활용
  - 이진 분류 (binary classification) 데이터를 학습하고 추정하기 위해 sigmoid (logistic function) 사용





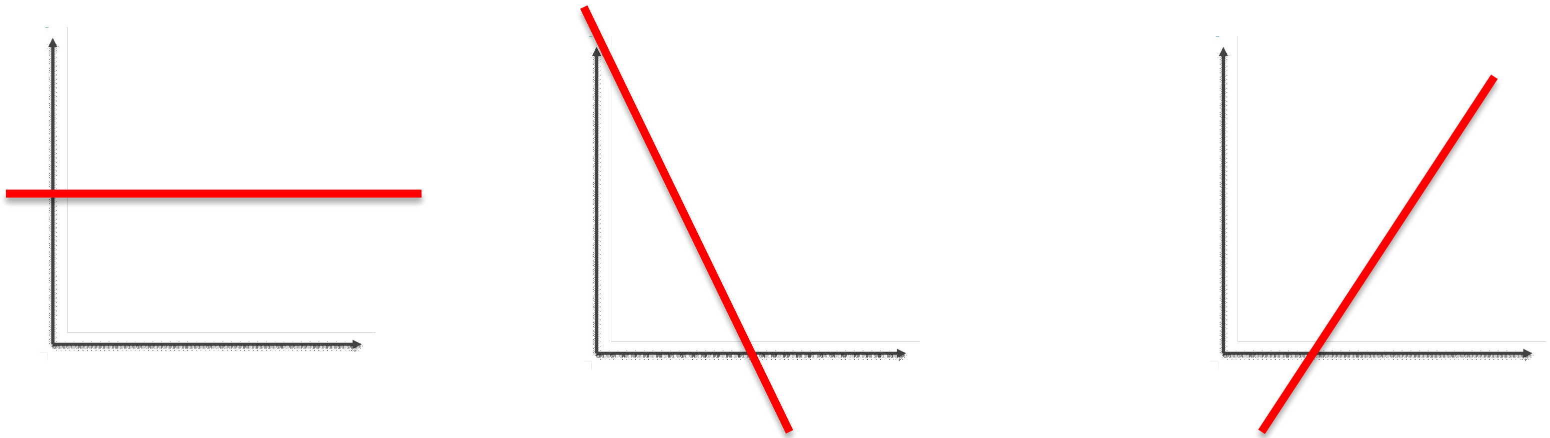
Logistic Regression for multiple classification



Multiclass Classification

❖ Method #1 : OvR (One vs Rest)

- 각 y에 대해 확률을 계산하고 확률이 가장 큰 y를 선택하자
- 각 모델을 독립적으로 학습



Multiclass Classification

❖ Method #2: Multinomial logistic regression

- 각 모델을 독립적으로 학습하는 방식이 아닌 동시에 학습 수행
- 각 모델을 학습하기 위해 one-hot encoding을 활용

no	x1	x2	y
1	0.8	1.2	0
2	9.8	2.0	2
3	3.2	3.5	0
4	4.8	5.2	1
...	...	...	...
N	9.2	2.5	2

⇨

	multiclass data			one-hot encoding		
no	x1	x2	y	p(y=0)	p(y=1)	p(y=2)
1	0.8	1.2	0	1	0	0
2	9.8	2.0	2	0	0	1
3	3.2	3.5	0	1	0	0
4	4.8	5.2	1	0	1	0
...	...	...	...	...	...	...
N	9.2	2.5	2	0	0	1

One-hot encoding

- a process used to convert categorical data variables into a form

y
1
1
3
2

1	0	0
1	0	0
0	0	1
0	1	0

Multinomial logistic regression

	multiclass data			one-hot encoding		
no	x1	x2	y	p(y=0)	p(y=1)	p(y=2)
1	0.8	1.2	0	1	0	0
2	9.8	2.0	2	0	0	1
3	3.2	3.5	0	1	0	0
4	4.8	5.2	1	0	1	0
...	...	...	...	...	...	...
N	9.2	2.5	2	0	0	1

x1	x2	p(y=0)
0.8	1.2	1
9.8	2.0	0
3.2	3.5	1
4.8	5.2	0
...	...	...
9.2	2.5	0

x1	x2	p(y=1)
0.8	1.2	0
9.8	2.0	0
3.2	3.5	0
4.8	5.2	1
...	...	...
9.2	2.5	0

x1	x2	p(y=2)
0.8	1.2	0
9.8	2.0	1
3.2	3.5	0
4.8	5.2	0
...	...	...
9.2	2.5	1

Multinomial logistic regression

$$log\frac{p(y=1)}{p(y=0)}=w_1x+b_1$$

$$\frac{p(y=1)}{p(y=0)}=exp(w_1x+b_1)$$

$$log\frac{p(y=2)}{p(y=0)}=w_2x+b_2$$

$$\frac{p(y=2)}{p(y=0)}=exp(w_2x+b_2)$$

$$p(y=1)=\frac{p(y=1)}{1}=\frac{p(y=1)}{p(y=0)+p(y=1)+p(y=2)}$$

$$=\frac{\frac{p(y=1)}{p(y=0)}}{\frac{p(y=0)}{p(y=0)}+\frac{p(y=1)}{p(y=0)}+\frac{p(y=2)}{p(y=0)}}$$

$$=\frac{exp(w_1x+b_1)}{1+exp(w_1x+b_1)+exp(w_2x+b_2)}$$

$w_0=0,\ b_0=0$

$$p(y=1)=\frac{\exp(w_1\cdot x+b_1)}{1+\exp(w_1\cdot x+b_1)+\exp(w_2\cdot x+b_2)} \leftarrow \frac{\exp(w_1\cdot x+b_1)}{\exp(w_0\cdot x+b_0)+\exp(w_1\cdot x+b_1)+\exp(w_2\cdot x+b_2)}=\frac{\exp(w_1\cdot x+b_1)}{\sum_{k=0}^{C-1}\exp(w_k\cdot x+b_k)}$$

softmax

(C : class 개수)

$$p(y=2)=\frac{\exp(w_2\cdot x+b_2)}{1+\exp(w_1\cdot x+b_1)+\exp(w_2\cdot x+b_2)}=\frac{\exp(w_2\cdot x+b_2)}{\sum_{k=0}^{C-1}\exp(w_k\cdot x+b_k)}$$

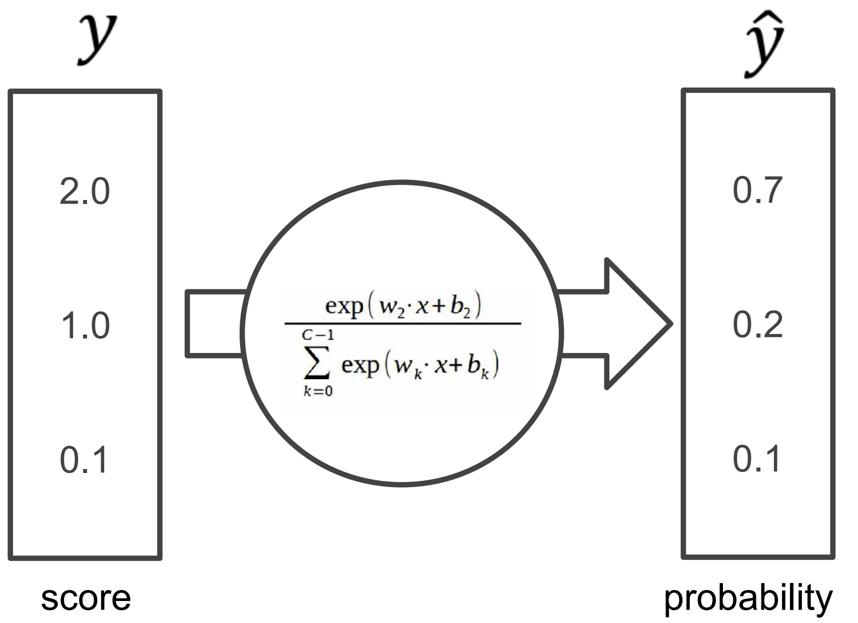
$$p(y=0)=\frac{1}{1+\exp(w_1\cdot x+b_1)+\exp(w_2\cdot x+b_2)}=\frac{\exp(w_0\cdot x+b_0)}{\sum_{k=0}^{C-1}\exp(w_k\cdot x+b_k)}$$

$$\Rightarrow \frac{\exp(w_2\cdot x+b_2)}{\sum_{k=0}^{C-1}\exp(w_k\cdot x+b_k)}$$

OvR은 각 model이 독립적으로 학습되지만, softmax는 한개 모델로 w_1, w_2, w_3, b_1, b_2, b_3를 동시에 학습한다.

Softmax

❖ 결과를 확률로 해석할 수 있게 변환해주는 함수



Loss Function (Cross Entropy)

y	$\hat{y}$				
	k= 0	1	2		
1	0.1	0.7	0.2	→ $0.1^0 \times 0.7^1 \times 0.2^0 = 0.7$	이 값이 클수록 추정을 잘한 것이다.
1	0.8	0.1	0.1	→ $0.8^0 \times 0.1^1 \times 0.1^0 = 0.1$	
0	0.8	0.1	0.1	→ $0.8^1 \times 0.1^0 \times 0.1^0 = 0.8$	

$$\boxed{P(y|x) = \prod_{k=0}^{C-1} \hat{y}_{i,k}^{\,1(y_i=k)}}$$

⇨ Likelihood

$$\hat{y}_{i,k} = \frac{\exp(w_k \cdot x_i + b_k)}{\sum_{k=0}^{C-1} \exp(w_k \cdot x_i + b_k)}$$ (i : data index, k : class index)

$$\boxed{L(w,b) = \prod_{k=0}^{C-1} \hat{y}_{i,k}^{\,1(y_i=k)}}$$

$1(y_i=k)$: indicator function
1(true) = 1
1(false) = 0

$$\log(L(w,b)) = \sum_{k=0}^{C-1} \log\left(\hat{y}_{i,k}^{\,1(y_i=k)}\right)$$

$$\log(L(w,b)) = \sum_{k=0}^{C-1} 1(y_i=k) \cdot \log(\hat{y}_{i,k})$$: $1(y_i=k) \rightarrow y_k$

$$\boxed{J(w,b) = \sum_{i=0}^{N-1} \sum_{k=0}^{C-1} y_k \cdot \log(\hat{y}_{i,k})}$$

N : data 개수, C : class 개수
← binary cross entropy에 대한 일반적 표현임.
C = 2라면 binary cross entropy.

- 목표 함수 (min. cross entropy)

$$\max_{w,b} J(w,b)$$

$$= \min_{w,b} \sum_{i=0}^{N-1} \sum_{k=0}^{C-1} [-y_k \cdot \log(\hat{y}_{i,k})]$$

$$= \min_{w,b} \sum_{i} CE$$

- Regularization

$$\min_{w,b} \sum_{i} CE + \lambda \sum_{k=0}^{C-1} |w_k|$$ (Lasso)

$$\min_{w,b} \sum_{i} CE + \lambda \sum_{k=0}^{C-1} w_k^2$$ (Ridge)

- 비교: binary classification

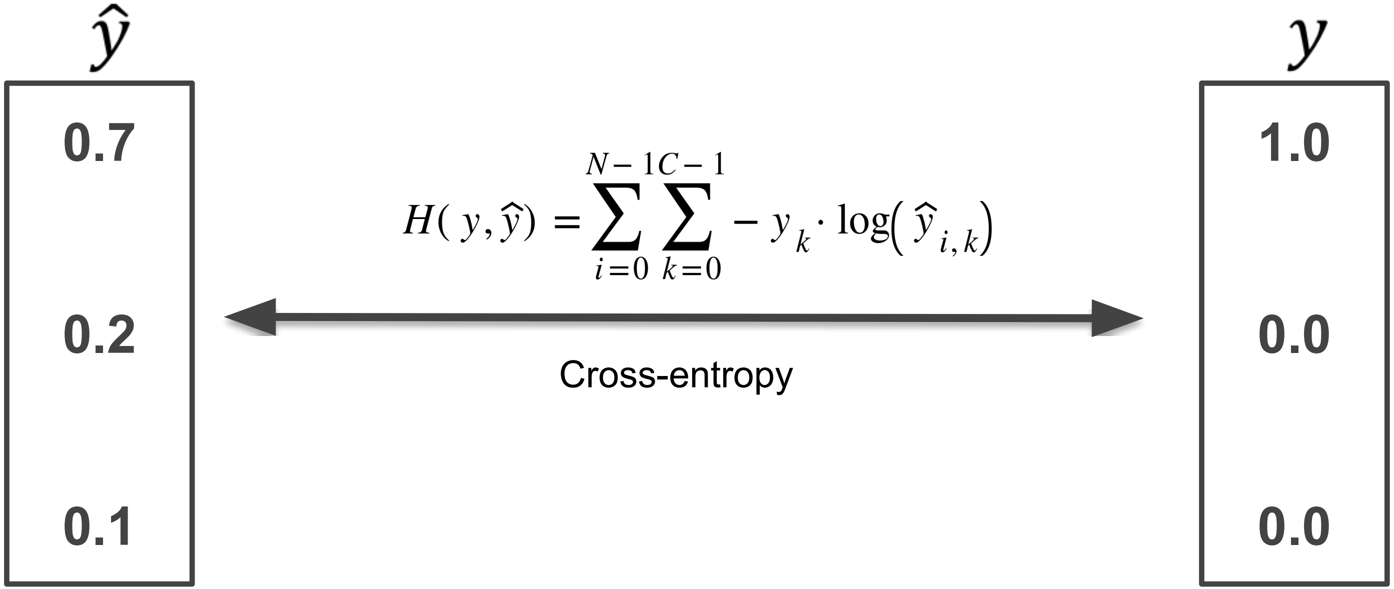
$$\hat{y}_i = \frac{1}{1 + \exp(-(wx_i + b))}$$

$$\min_{w,b} \sum_{i} BCE$$

Loss Function (Cross Entropy)

- Cross-entropy

$$H(y,\hat{y}) = \sum_{i=0}^{N-1}\sum_{k=0}^{C-1} -y_k \cdot \log\left(\hat{y}_{i,k}\right)$$



Multinomial logistic regression

$$p_{n,k} = P(Y = k | X = \mathrm{x}_n) = \frac{\exp(z_{n,k})}{\sum_j \exp(z_{n,j})}$$

Softmax function

x_n

Probabilities must be >= 0

Probabilities must sum to 1

y_n

	x_n		Unnormalized Probabilities		Probabilities				Correct prob.
cat	**3.2**		**24.5**		**0.13**				**1.00**
car	5.1	$\xrightarrow{\text{exp}}$	164.0	$\xrightarrow{\text{norm.}}$	0.87	$\longrightarrow$	**Compare**	$\longleftarrow$	0.00
frog	-1.7		0.18		0.00				0.00

softmax

Cross Entropy