

03/03 지도학습 [ Regression (회귀) → 연속적인 숫자값을 prediction ] → Linear Regression (선형회귀)

[ Classification (분류) → 입력값이 0 or 1 중 어디에 속하는가 (binary classification 아닌 분류) ] → Logistic Regression

- KNN
- Decision Tree
- Naive Bayes
- ...

• Linear Regression (Hypothesis (predict model)) →  $w x + b$

loss function →  $E(w, b) = \frac{1}{n} \sum_{i=1}^n [t_i - (w x_i + b)]^2$

• Logistic Regression (Hypothesis  $y = \frac{1}{1 + e^{-(w x + b)}}$ )

sigmoid →  $\sigma(x) = \frac{1}{1 + e^{-x}}$

loss →  $E(w, b) = \frac{1}{n} \sum_{i=1}^n [t_i - (\frac{1}{1 + e^{-(w x_i + b)}})]^2$

이렇게 만들면 convex 함수가 X

• Logistic Regression

$$\begin{cases} \text{Hypothesis: } y = \frac{1}{1 + e^{-(wx+b)}} & \text{← 예측 리미트는 0 or 1} \\ \text{loss: } & \end{cases}$$

$$\left( \text{Cross Entropy} \right) = - \sum_{i=1}^n \{ x_i \log y_i + (1-x_i) \log (1-y_i) \} \quad \leftarrow \text{convex}$$

log loss

→ 가짜 문제를 공부 구현해 보아요!

$\begin{bmatrix} \bullet \text{ GRE} \rightarrow \\ \bullet \text{ GPA} \rightarrow \\ \bullet \text{ RANK} \rightarrow \end{bmatrix} \rightarrow \text{특성을 사용} \Rightarrow \text{prediction}$

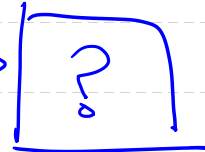
★ Python 구현  
 ★ Tensorflow 구현  
 ★ Sklearn 구현

구현한 후

$\begin{bmatrix} \text{GRE} \rightarrow 600 \\ \text{GPA} \rightarrow 3.8 \\ \text{RANK} \rightarrow 1 \end{bmatrix}$

→ 이 경우에 대한 prediction

결과??



이진분류 (binary Classification) (Metrics)

↓ "logistic Regression" → 분류 성능 평가 지표

우리가 원하는 logistic Regression model이 잘 만들어진 모델인지 확인하기 위한 기준

## Confusion Matrix

② Recall (재현율)  
→ 실제 True인 것중에 우리 Model이 True로 예측한 것의 비율.

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{sensitivity hit rate}$$

Confusion Matrix	실제정답	
	True (1)	False (0)
분류결과	True positive (TP)	False (FP)
	False Negative (FN)	True Negative (TN)

① Precision (정밀도)

→ 우리가 모델이 True로 분류한 것중에 실제로 True인 것의 비율

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Positive Predictive Value (PPV)}$$

Model		
Actual	Predict	Predict Value
1	1	→ x
2	0	→ x
3	1	→ ok
4	0	→ ok
5	?	← Prediction

· A라는 쿼리가 왔어요!!

소프트웨어를 개발 → 사진을 입력받아서 그 안에 고양이인지를 검색하는

✓ 호보 → 이 사진의 고양이 검색률이 99.9%예요. ] → 리사의 프레를 갖고 생각하는거  
[ 이 사진의 " " 63.7% " " ] "이 사진은 오검"을 하지 않는다"

[ (A) "A" 사진 99.9%의 검색률 + 그림 상동 평균 5개의 오검. ]  
✓ (B) "B" 사진 63.7%의 검색률 + 오검이 거의 X ✓

(정확도)  
정확도  
"Precision"

Recall (재현율)  
↓  
검색률

★ ( Recall : 고양이를 빠뜨리지 않고 얼마나 잘 검색해 내는가 라는 지표 ] → 반비례 ( )  
Precision : 검색된 결과가 얼마나 정확하게 나타내는지 지표.

Precision - Recall graph!

• Accuracy (정확도) (Precision, Recall)  $\rightarrow$  True를 True로 옳게 예측한 경우

$\hookrightarrow$  False를 False로 예측하는 경우도 포함!

$\Delta$  accuracy (정확도) =  $\frac{\textcircled{TP} + \textcircled{TN}}{TP + FN + FP + TN}$   $\Rightarrow$  accuracy 계산

"domain의 bias를 반드시 고려"