

# Chapter3.Classification

신준호

# 목표

- 1. MNIST 예제를 통해 분류기법 이해
- 2. 어떤 분류모델이 있고,  
어떻게 정확하게 분류할 것인지 학습.

# 목차

1. Classifier 가 무엇이고 어떻게 있는지.

- Binary Classifier
- Multi Class Classification
- Multi Label Classification
- Multi Output Classification

2. Performance Measure를 하는 방법.

- Cross Validation
- Confusion Matrix
- Precision and Recall Trade-off
- F1 Score
- ROC Curve
- Error Analysis

# Classifier

- Classifier란 ?  
: 인풋에 대한 예측값을 어떻게 분류하는 시스템
- Classifier 종류
  - Binary Classifier
  - Multi Class Classification
  - Multi Label Classification
  - Multi Output Classification

# Classifier

- Binary Classifier 란 ?
  - : 인풋에 대한 예측값이 ~인지 혹은 ~아닌지 두 가지의 클래스로 분류하는 방법
  - Ex) 인풋 이미지가 5인지, 5가 아닌지.

# Classifier

- Multi Class Classification 란 ?  
: Binary와 다르게 두 개 이상의 클래스로 분류를 한다.

## 방법

- OvA(One-Versus-All)  
: 인풋에 대한 모든 경우의 클래스를 분류하는  
Ex) 1이라는 이미지가 1~10의 숫자를 분류하는 10개의 클래스로 분류하는 방법
- OvO(One-Versus-One)  
: 인풋에 대한 구분하고자 하는 두 개의 클래스로 분류  
Ex) 1이라는 이미지가 1인지, 2인지 두 개의 클래스로 분류하는 방법
-

# Classifier

- Multi Label Classification 란 ?

: 인풋에 대한 예측값이 하나의 클래스에만 할당되는게 아니라 여러 클래스로 출력하도록 하는 분류

Ex) 단체사진의 인풋 인스턴스에 각 인물들의 Label을 붙여 주는 것

# Classifier

- Multi Output Classification 란 ?  
: 각 라벨이 Multiclass 를 가질 수 있는 Multi Label Classification의 일반화이다.  
라벨이 2개 이상의 값을 가질 수 있다.  
(값에 Range가 있을 경우 Range 범위안에서 값을 가짐.)

Ex) Noise removing System이 있다면 인풋 이미지에서 Noise를 없앤 깨끗한 이미지로 출력을 한다.  
각 라벨은 이미지의 1Pixel & Pixel value Range( 0 ~ 255)



# Performance Measure

- Performance Measure 란 ?
  - : Classifier의 예측 정확도를 측정하는 것
  - 더 좋은 Classifier를 만들기 위한 것

방법

Cross Validation

Confusion Matrix

Precision and Recall Trade-off

F1 Score

ROC Curve

Error Analysis

# Performance Measure

- Cross Validation방법

: 주로 K-Fold Cross validation 을 활용해 Train Dataset 을 K개의 Folder로 나눠놓고 모델을 학습시키고 평가하는 방법

# Performance Measure

- Confusion Matrix 방법

: True / False는 실제 값이 1인지 0인지를 맞췄는지, 못 맞췄는지.

Positive / Negative 예측한 값이 맞았는지 안 맞았는지.

TP: 예측한 값이 5, 실제로 5

TN: 예측한 값이 X, 실제로 X

FP: 예측한 값이 5, 실제로 X

FN: 예측한 값이 X, 실제로 5

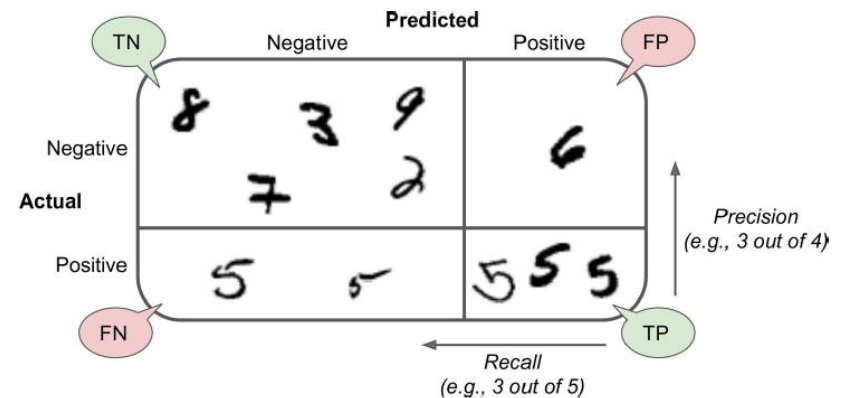


Figure 3-2. An illustrated confusion matrix

# Performance Measure

- Precision and Recall Trade-off 방법

: 정확도(Precision)란 맞다고 예측한게 실제로 맞았는지의 지표  
, 재현율(Recall)란 실제 정답의 True중 얼마나 많은 예측한게 실제로 맞았는지의 지표

Trade-off : Precision 과 Recall의 임계값(ThreshHold)을 조정해주는 역할

F1 Score : 조화평균 ( Precision과 Recall을 하나로 결합)

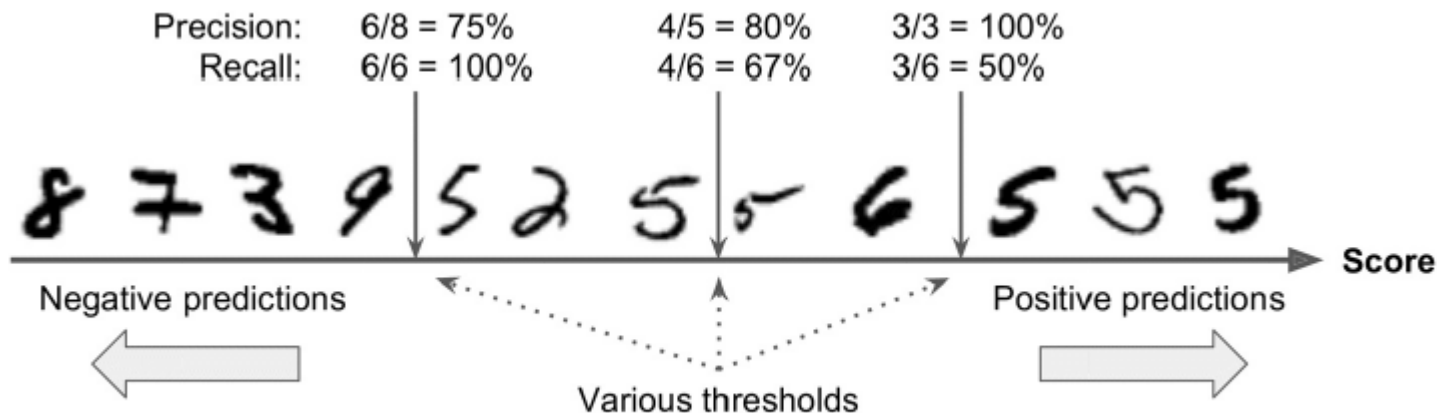


Figure 3-3. Decision threshold and precision/recall tradeoff

# Performance Measure

- Precision and Recall Trade-off 방법

: 정확도(Precision)란 맞다고 예측한게 실제로 맞았는지의 지표  
, 재현율(Recall)란 실제 정답의 True중 얼마나 많은 예측한게 실제로 맞았는지의 지표

Trade-off : Precision 과 Recall의 임계값(ThreshHold)을 조정해주는 역할

F1 Score : 조화평균 ( Precision과 Recall을 하나로 결합)

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

Equation 3-3.  $F_1$  score

$$F_1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{TP}{TP + \frac{FN + FP}{2}}$$

# Performance Measure

- Precision/Recall Curve 방법  
: precision/recall trade-off 의 적절한 Threshold 값을  
찾기위한 분석 방법

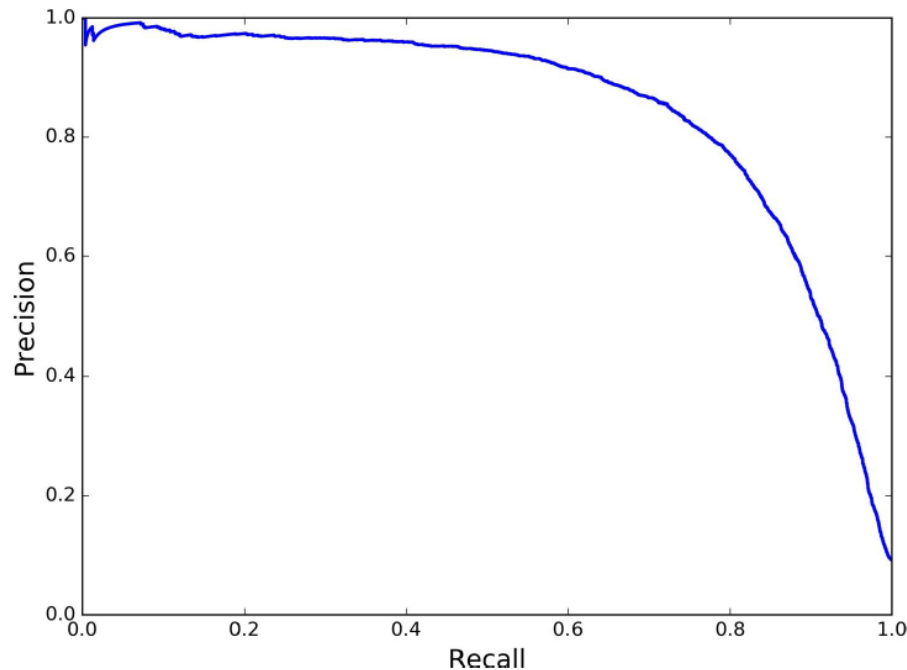


Figure 3-5. Precision versus recall

# Performance Measure

- ROC Curve 방법

: Receiver Operation Characteristic Curve의 약자로

*True Positive rate* (another name for recall) against the *False Positive rate*

$$FPR = 1 - TNR \text{ (specificity)}$$

Hence the ROC curve plots *sensitivity* (recall) versus  $1 - \textit{specificity}$ .

# Performance Measure

- ROC Curve 방법

민감도(Sensitivity) : 5인 케이스에 대해 5이라고 예측

특이도(Specificity) : X인 케이스에 대해 X이라고 예측

-민감도 = True Positive = Recall

-특이도 = True Negative

-False Positive =  $1 - \text{특이도} = 1 - \text{TNR}$



# Performance Measure

- ROC Curve 방법

$TPR$  ( recall *sensitivity* )

$FPR = 1 - TNR$  ( *specificity* )

ROC curve plots

*sensitivity* (recall) versus  $1 - \textit{specificity}$ .

잘 분류된 모델일 수록  
Top Left 쪽으로 간다.

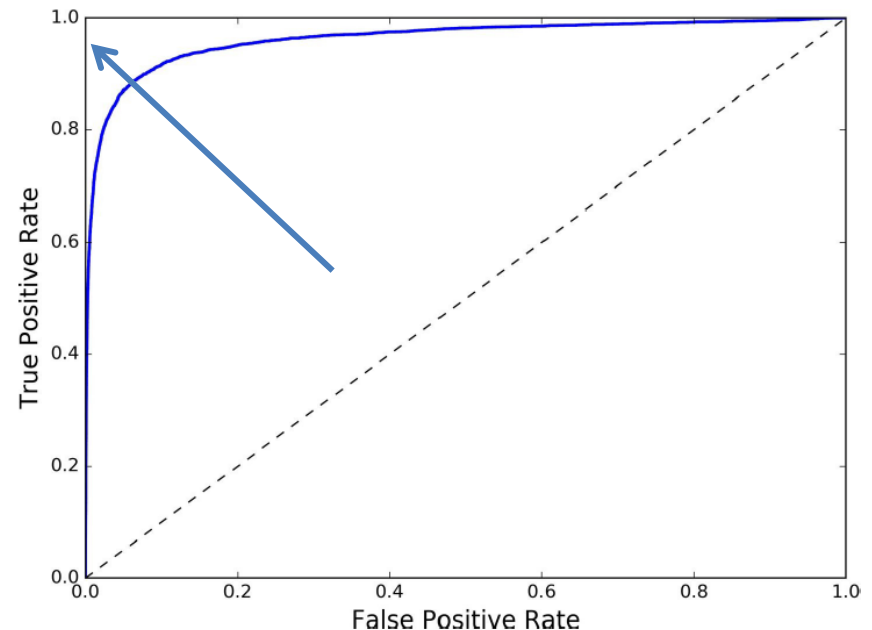


Figure 3-6. ROC curve

# Performance Measure

- ROC Curve 방법

ROC Curve와 PR Curve가 유사해서 언제 어떤걸 써야할지 궁금할 수 있는데. Positive class가 드물거나 False Negative보다 False Positive에 대해 유의해야 되는 상황이면 Precision/Recall 커브를 사용해야 한다.

그렇지 않으면 ROC Curve를 사용한다.

Fp: 5가 아닌데 5라고 예측

TP: 5인데 5라고 예측

Fn: 5인데 5가 아니라고 예측한거

TN: 5가 아닌데 5가 아니라고 예측한거

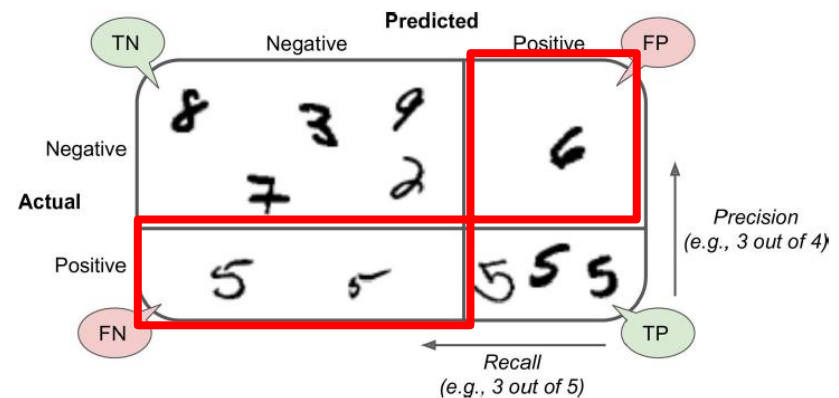


Figure 3-2. An illustrated confusion matrix

# Performance Measure

- Error Analysis 방법  
: 모델의 성능을 향상시키는 방법으로 오류유형을 분석.

Confusion Matrix를 확인 하면서

각 값의 해당 클래스 별 에러율을 확인하고 데이터를 확인  
해 오류를 줄일 수 있는 방법을 찾아내는 것이다.