



ML/DS 지식 따라가기

Evaluation Techniques



EVALUATION

머신러닝 모델을 만들었으면, 잘 돌아가는지 확인을 해야함

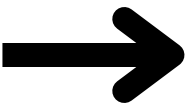


EVALUATION

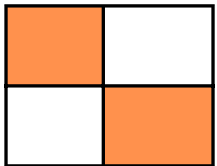
evaluation의 종류 - 분류

Confusion Matrix

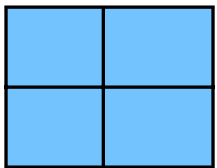
		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN



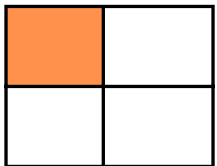
Accuracy
(정확도)



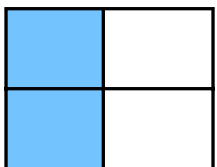
—



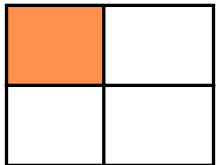
Precision
(정밀도)



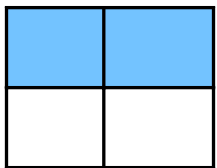
—



Recall
(재현도)



—



EVALUATION

evaluation의 종류 - 분류

Confusion Matrix

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

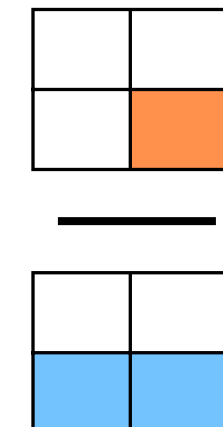


F1 Score

$$\frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

TPR = Recall

FPR = 1 - TPR =

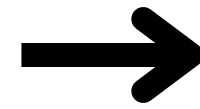


EVALUATION

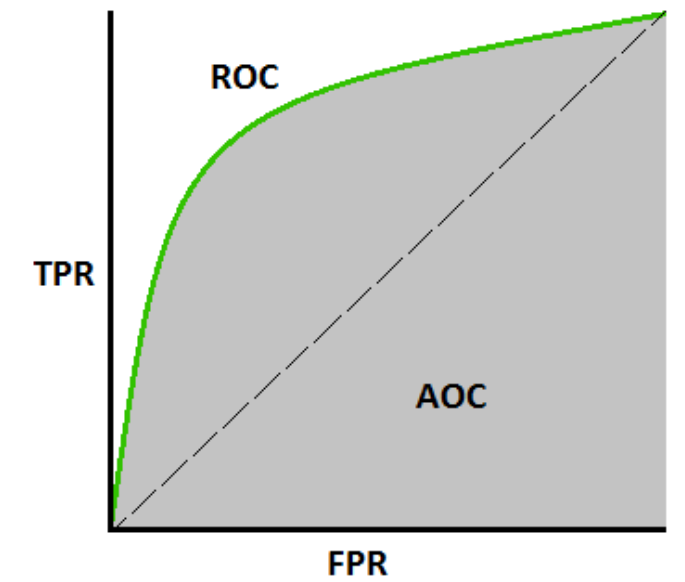
evaluation의 종류 - 분류

Confusion Matrix

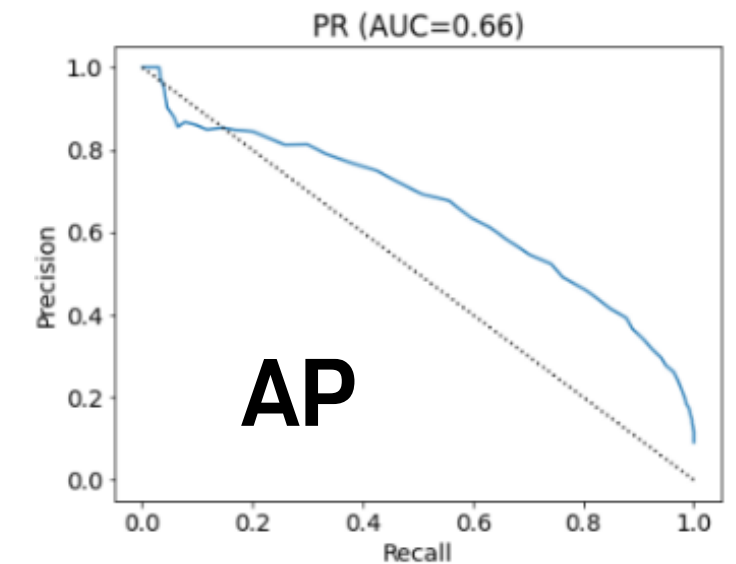
		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN



ROC X축 : FPR
Y축 : TPR



PR X축 : Recall
Y축 : Precision

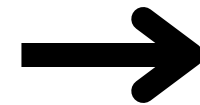


EVALUATION

evaluation의 종류 - 분류

Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN



Matthews Correlation Coefficient

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

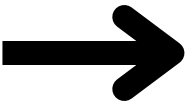
데이터가 불균형해서 accuracy 값이 이상한 경우에 사용함

EVALUATION

evaluation의 종류 - 분류

Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN



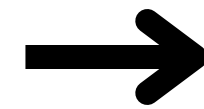
		Predicted			
		A	B	...	C
Actual	A	24	24	...	35
	B	57	75	...	46
	68
	C	86	57	68	85

EVALUATION

evaluation의 종류 - 분류

Confusion Matrix

		Predicted			
		A	B	...	C
Actual	A	24	24	...	35
	B	57	75	...	46
	68
	C	86	57	68	85



Accuracy
(정확도)

$$\frac{\text{맞은 개수}}{\text{전체}}$$

EVALUATION

evaluation의 종류 - 분류

Confusion Matrix

		Predicted			
		A	B	...	C
Actual	A	24	24	...	35
	B	57	75	...	46
	68
	C	86	57	68	85



**Micro Precision
(마이크로 정밀도)**

분류도 됐고 맞기도 한 것

해당 클래스로 분류된 것

**Macro Precision
(매크로 정밀도)**

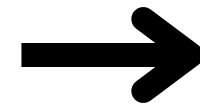
각 클래스의 Micro Precision의
평균을 구한 것

EVALUATION

evaluation의 종류 - 분류

Confusion Matrix

		Predicted			
		A	B	...	C
Actual	A	24	24	...	35
	B	57	75	...	46
	68
	C	86	57	68	85



Micro Recall
(마이크로 재현율)

분류도 됐고 맞기도 한 것

정답이 해당 클래스인 것

Macro Recall
(매크로 재현율)

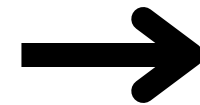
각 클래스의 Micro Recall의
평균을 구한 것

EVALUATION

evaluation의 종류 - 분류

Confusion Matrix

		Predicted			
		A	B	...	C
Actual	A	24	24	...	35
	B	57	75	...	46
	68
	C	86	57	68	85



Micro F1
(마이크로 F1)

Macro F1
(매크로 F1)

$2 * \text{Micro Precision} * \text{Micro Recall}$

$\text{Micro Precision} + \text{Micro Recall}$

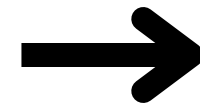
각 클래스의 Micro F1의
평균을 구한 것

EVALUATION

evaluation의 종류 - 분류

Confusion Matrix

		Predicted			
		A	B	...	C
Actual	A	24	24	...	35
	B	57	75	...	46
	68
	C	86	57	68	85



Micro-average F1
(마이크로-평균 F1)

클래스별 Binary CM을 더한 후
그것을 기준으로 F1을 구한 것

Weighted F1
(가중 F1)

F1 식에 데이터 수로
가중치를 곱해줌

EVALUATION

evaluation의 종류 - 분류

Confusion Matrix

		Predicted			
		A	B	...	C
Actual	A	24	24	...	35
	B	57	75	...	46
	68
	C	86	57	68	85

Kohen's Kappa

$$k = \frac{p_0 - p_e}{1 - p_e} = 1 - \frac{1 - p_0}{1 - p_e}$$

p0 : accuracy

pe : 우연히 맞혔을 확률

= (대각선에 있는 값 / 전체 개수)의 평균

EVALUATION

evaluation의 종류 - 회귀

Error값들

- Mean of Absolute Error
(평균 절대 오차)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - f(\mathbf{x}_i)|$$

특징

1. 특이값이 많을 때 사용함.
2. 값이 낮을수록 좋음

장점

1. 직관적임.
2. 정답값과 예측값이 같은 단위임.

단점

1. 정답값을 상회하는지 못 미치는지
를 알 수 없음.
2. 크기에 의존적임

EVALUATION

evaluation의 종류 - 회귀

Error값들

- Mean of Squared Error
(평균 제곱 오차)

$$MSE = \frac{\sum_{i=1}^n (y - \hat{y})^2}{n}$$

특징

1. 값이 낮을수록 좋음

장점

1. 직관적임.

단점

1. 1 미만은 작아지고, 그 이상은 커짐.
2. 정답값을 상회하는지 못 미치는지 알 수 없음.
3. 크기에 의존적임

EVALUATION

evaluation의 종류 - 회귀

Error값들

- Root Mean of Squared Error
(평균 제곱근 오차)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y - \hat{y})^2}{n}}$$

특징

1. 값이 낮을수록 좋음
2. 루트를 씌워서 값의 왜곡을 줄임

장점

1. 직관적임.

단점

1. 1 미만은 작아지고, 그 이상은 커짐.
2. 정답값을 상회하는지 못 미치는지 알 수 없음.
3. 크기에 의존적임

EVALUATION

evaluation의 종류 - 회귀

Error값들

- Mean Absolute Percentage Error
(평균 절대 비율 오차)

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y - \hat{y}}{y} \right|$$

특징

1. 크기 의존적인 문제를 해결함.
2. 값이 낮을수록 좋음

장점

1. 직관적임.
2. 다른 모델들과 에러율 비교가 쉬움.

단점

1. 정답값을 상회하는지 못 미치는지 알 수 없음.
2. 정답값이 1보다 작으면 값이 무한대로 수렴할 수 있음

EVALUATION

evaluation의 종류 - 회귀

Error값들

- **Mean Percentage Error**
(평균 비율 오차)

$$MPE = \frac{100}{n} \sum_{i=1}^n \frac{y - \hat{y}}{y}$$

특징

1. MAPE에서 절대값을 뺀

장점

1. 정답값을 상회하는지 못 미치는지
를 알 수 있음.
2. 편향을 측정할 수 있음

단점

1. 에러가 일부 상쇄될 수 있음.
2. 실제 값이 0이면 사용 불가능함.

EVALUATION

evaluation의 종류 - 회귀

Error값들

- **Symmetric Mean of Absolute Percentage Error**
(대칭 평균 절대 비율 오차)

$$SMAPE = \frac{100}{n} \times \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{(|Y_i| + |\hat{Y}_i|) / 2}$$

특징

1. MAPE이 가지는 한계점을 보완함

장점

1. 0~200%의 값을 가지므로 해석하기 더 좋음
2. 비율을 기반으로 한 값이므로 모델 간 성능 비교를 하기 좋음
3. 실제 값에 0이 있어도 계산이 가능함.

단점

1. 실제 값과 예측 값이 전부 0이면 계산이 불가능함.
2. 둘 중 하나가 0이며도 값이 매우 커짐.

EVALUATION

evaluation의 종류 - 회귀

Error값들

- R² (결정계수)

$$R^2 = 1 - \frac{SSE}{SST}$$

특징

1. 상대적인 성능을 보여주기 좋음.
2. 1에 가까울수록 좋음
3. "설명력"이라는 개념을 사용함.

장점

1. 데이터 크기의 영향을 받지 않음.
2. 이상치에 덜 민감함.

= MSE

$$SSE : \frac{1}{n} \sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2$$

$$SST : \frac{1}{n} \sum_{i=1}^n (y^{(i)} - \mu_y)^2$$

= 분산

단점

1. feature가 많아질수록 같이 올라감.

EVALUATION

evaluation의 종류 - 회귀

Error값들

- adjusted R^2 (수정된 결정계수)

$$R^2 = 1 - \frac{SSE}{SST}$$



$$adj.R^2 = 1 - \frac{(n-1)}{(n-p)} (1 - R^2)$$

EVALUATION

evaluation의 종류 - 군집

Internal Evaluation

- Silhouette Score (실루엣 점수)

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

a : 자신이 속한 군집의 평균 거리
b : 다른 군집과의 최소 거리

장점

- 클러스터 개수를 판단하기 좋음
- 시각화하기 좋음

단점

- 데이터 개수가 많아지면 계산량도 엄청 많아진다.

EVALUATION

evaluation의 종류 - 군집

Internal Evaluation

- **Dunn Index (던 지수)**

$$I(C) = \frac{\min_{i \neq j} \{d_c(C_i, C_j)\}}{\max_{1 \leq l \leq k} \{\Delta(C_l)\}}$$

(가장 작은 군집 간 거리) / (가장 큰 군집 내 거리)

장점

- 간단하고 직관적임

단점

- 클러스터 수에 영향을 받음
- 데이터 크기에 영향을 받음

EVALUATION

evaluation의 종류 - 군집

External Evaluation

- Rand Index (랜드 지수)

$$RI = \frac{a + b}{\binom{n}{2}}$$

(정답인 쌍) / (가능한 모든 쌍)



$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]}$$

adjusted Rand index

EVALUATION

evaluation의 종류 - 군집

External Evaluation

- Fowlkes-Mallows Index (파울쿠스-말로우 지수)

$$FM = \sqrt{\frac{TP}{TP + FP} \cdot \frac{TP}{TP + FN}}$$

군집 간 유사성을 측정함

EVALUATION

evaluation의 종류 - 군집

External Evaluation

- Fowlkes-Mallows Index (파울쿠스-말로우 지수)

$$FM = \sqrt{\frac{TP}{TP + FP} \cdot \frac{TP}{TP + FN}}$$

군집 간 유사성을 측정함

EVALUATION

Thank you