

# モデル評価指標

ークラス分類編ー

株式会社KSKアナリティクス



# 目次

---

- はじめに
- 分類モデルの評価  
Performanceオペレータ
- 分類モデルの評価指標  
代表的な評価指標
- おわりに

# はじめに

## 背景

機械学習モデルの作成において、  
良いモデルかどうかを判断するためには  
重視する評価指標を決めて判断する必要があります。

しかし、どの評価指標をみて判断するのが良いか、  
RapidMinerのPerformanceオペレータは種類が多く  
どれを使うか迷ってしまいます。  
これらの悩みを解決するため、評価指標をまとめました。

今回は、分類モデルの評価指標についてご紹介します。

# 分類モデルの評価

## Performanceオペレータ

目的変数が2値または3値以上かによって選択します。

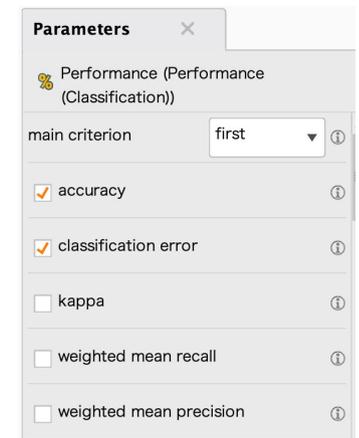
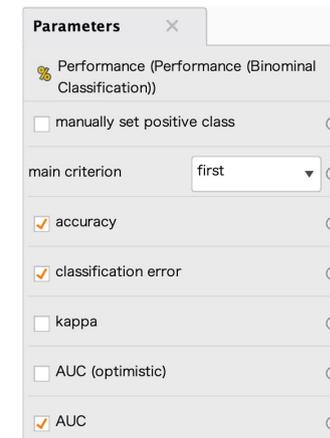
### 2値分類

Performance(Binominal Classification)



### 3値以上

Performance(Classification)



<補足>

Performance (記載なし)

分類か回帰など自動的に決定し、最も一般的な評価基準を計算します。

より高度な評価をするためには、上記オペレータの選択をおすすめします。

# 分類モデルの評価

## Performanceパラメータ

### パラメータ設定方法

- 1 2で選択する指標のうち重視する指標を選択  
“first”の場合、並びが最初にくる指標になる
- 2 結果画面に表示する指標を選択

では、どの評価指標を選択すれば良いか。  
代表的な評価指標を2値分類モデルを例にご紹介します。

Parameters

Performance (Performance (Binominal Classification))

manually set positive class

1 main criterion first

2  accuracy

classification error

kappa

AUC (optimistic)

AUC

AUC (pessimistic)

precision

recall

lift

fallout

f measure

# 分類モデルの評価指標

## 代表的な評価指標

- 混同行列 (Confusion Matrix)
- Accuracy (正答率)
- Classification Error (分類誤差)
- Precision (適合率)
- Recall (再現率)
- F-measure (F値) ※不均衡データ
- AUCとROC曲線

# 分類モデルの評価指標

## 混同行列 (Confusion Matrix)

### 教師あり分類の予測結果と実際の値の違いをみる

混同行列は、クラス分類結果をまとめた表です。

		実際	
		true A(Positive)	true B(Negative)
予測	pred. A (Positive)	TP	FP
	pred. B (Negative)	FN	TN

文字の意味 T(True):予測正解、F(False):予測が不正解、P(Positive):正と予測、N(Negative):負と予測

# 分類モデルの評価指標

## 混同行列 (Confusion Matrix)

### 具体例1 : ウイルスPCR検査

感染の疑いがある人に対してPCR検査を実施し陰性または陽性の診断を混同行列(右図)で表します。陰性患者を陽性と誤診するよりも、陽性患者を陰性と誤診してしまう方が一大事に繋がりがねません。この場合は、陽性患者を見逃さずどれだけ検出できたかの指標を選択するのが適切です。

	実際は陽性	実際は陰性
予測が陽性	TP 陽性の人に陽性と診断	FP 陰性の人に陽性と診断 誤り
予測が陰性	FN 陽性の人に陰性と診断 誤り	TN 陰性の人に陰性と診断

### 具体例2 : 品質検査

正常と予測し実際は異常であった時の見逃しはクレームや問題に繋がります。この場合は、異常を見逃さずどれだけ検出できたかの指標を選択します。正常を異常予測し無駄になるリスクを避け、異常予測の確実性を重視する場合は、異常予測がどれだけ正解しているかの指標を選択します。

➡ **何を重視するかによって選択する指標が変わるため、目的に応じて評価指標を選択することが重要になります。**

# 分類モデルの評価指標

## 混同行列 (Confusion Matrix)

混同行列を使って評価指標を算出することができます。

### Accuracy (正答率)

$$=(TP+TN)/(TP+TN+FP+FN)$$

### Precision (適合率)

$$=TP/(TP+FP)$$

### Recall (再現率)

$$=TP/(TP+FN)$$

### F-measure (F値)

=適合率と再現率の調和平均

		実際	
		true A (Positive)	true B (Negative)
予測	pred. A (Positive)	TP	FP
	pred. B (Negative)	FN	TN

positive classがAの場合

# 分類モデルの評価指標

Accuracy (正答率) / Classification Error (分類誤差)

**メリット** : 全ての予測結果がわかる、シンプルでわかりやすい

**デメリット** : クラスのデータが不均衡の場合、信用できない指標になる

## ■ Accuracy (正答率)

全予測に対する正解の割合

何%以上であれば良いかは、業界や課題によって違います。

例えば、出荷検査は99.9%を求める場合や、マーケティング購入予測では65%以上あれば上出来の場合もあります。

## ■ Classification Error (分類誤差)

全予測に対する不正解の割合

accuracy: 83.89% +/- 3.68% (micro

pred. A

pred. B

交差検証した結果  
平均からのばらつきを示す  
(≒標準偏差)

# 分類モデルの評価指標

## Precision (適合率)

**メリット** : 誤検出を発見できる

**デメリット** : 見逃しを発見できない

### ■ Precision (適合率)

positive と予測したもののうち、どれだけ正解したか

precision: 79.77% +/- 7.86% (micro average: 79.25%) (positive class: B) 目的のクラスがpositive classに表示されているか確認できます。

	true A	true B	class precision
pred. A	512	70	87.97%
pred. B	66	252	79.25%
class recall	88.58%	78.26%	

### P.8 具体例 : 品質検査

正常を異常と予測し無駄になるリスクを避け、異常予測の確実性を重視する場合、異常予測の確実性を上げるため、適合率の指標を選択しモデルを調整します。

# 分類モデルの評価指標

## Recall (再現率)

**メリット** : 見逃しを発見できる

**デメリット** : 誤検出を発見できない

### ■ Recall (再現率)

実際にpositive であるものをどれだけ検出できたか

recall: 78.29% +/- 5.13% (micro average: 78.26%) (positive class: B)

	true A	true B	class precision
pred. A	512	70	87.97%
pred. B	66	252	79.25%
class recall	88.58%	78.26%	

### P.8 具体例 : 品質検査

正常と予測し実際は異常であった時の見逃しはクレームや問題に繋がります。異常を見逃さずに検出するため、再現率の指標を選択しモデルを調整します。

# 分類モデルの評価指標

## F-measure (F値)

### 見逃しと誤検出を均等に判断する

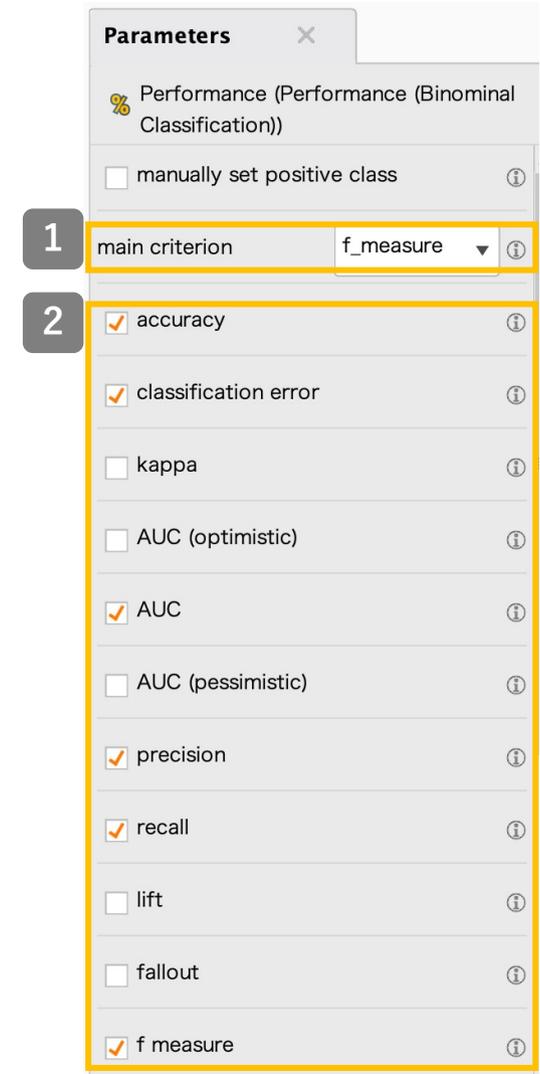
#### ■ F-measure (F値)

適合率と再現率の調和平均

適合率と再現率は**トレードオフの関係**にあります。  
クラスのデータに偏りがある場合（不均衡データ）  
はF値を選択しモデルを調整します。

#### パラメータ設定方法

- 1 “F-measure”を選択します。
- 2 評価指標を右図のように選択します。



# 分類モデルの評価指標

## AUCとROC曲線

### 間違いを許容するポイントをグラフでみる

#### ■ AUC (Area Under the Curve)

2値分類モデルの性能を測定、可視化

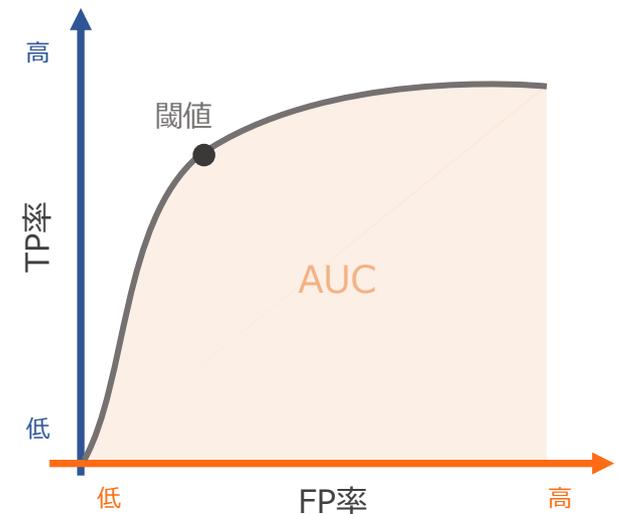
AUC(ROC曲線下の面積) が大きくなるように閾値を調整します。

#### □ ROC曲線

縦軸にTP率、横軸にFP率をとる曲線

TP率が大きくFP率が小さくなるように閾値を調整  
複数のモデルの測定を比較する場合は、ROC曲線の  
カーブが直角に近いほど良いモデルと判断できます。

TP率(true positive rate) 、FP率(false positive rate) : P.7 混同行列 参照



## まとめ

評価指標は、目的によって選択することが重要です。  
モデルの評価で誤った判断をしないために、評価指標の特徴や違いを理解し、より良いモデルを作成することで課題を解決していただきたいと考えております。

次回は、回帰モデルの評価指標についてご紹介します。