

ルールアンサンブルを用いた マイノリティクラスの識別

情報認識学研究室 修士2年

横山 祐也

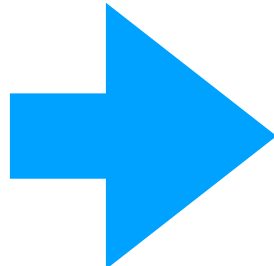
背景

パターン認識において度々、『クラスインバランス』が問題となっている.

例: 癌が陰性か陽性かの分類問題

データセット

クラス名	陰性	陽性
データ数	90	10

モデル		真陰性	真陽性		通常の精度
○	識別モデル1	81/90	1/10		82%
×	識別モデル2	63/90	7/10		70%

医療分野などでは, 識別モデル2が望ましい.

目的

1. データ数によるマジョリティクラスの影響をなるべく抑えて、マイノリティクラスの識別率の向上を図る.

	モデル	真陰性	真陽性
✕	識別モデル1	81/90	1/10
○	識別モデル2	63/90	7/10
✕	識別モデル3	1/90	10/10

2. 加えて, 識別モデルがマイノリティクラスと判断した根拠を知りたい.

➡ 希少疾患など

提案

1. 識別モデルでバランスされた精度(Balance Accuracy)の最大化を図る

	モデル	真陰性	真陽性	バランスされた精度
✕	識別モデル1	81/90	1/10	50%
○	識別モデル2	63/90	7/10	70%
✕	識別モデル3	1/90	10/10	51%

計算の例: 識別モデル2

$$\frac{1}{2} \cdot \left(\frac{63}{90} + \frac{7}{10} \right) \cdot 100 = 70\%$$

2. ルールアンサンブル法の採用

決定木から抽出したルールをベースにした線形回帰モデル.



決定ルールの重要度を測定することができる