

データサイエンス研究室

～データの背後にある確率構造をモデル化する～

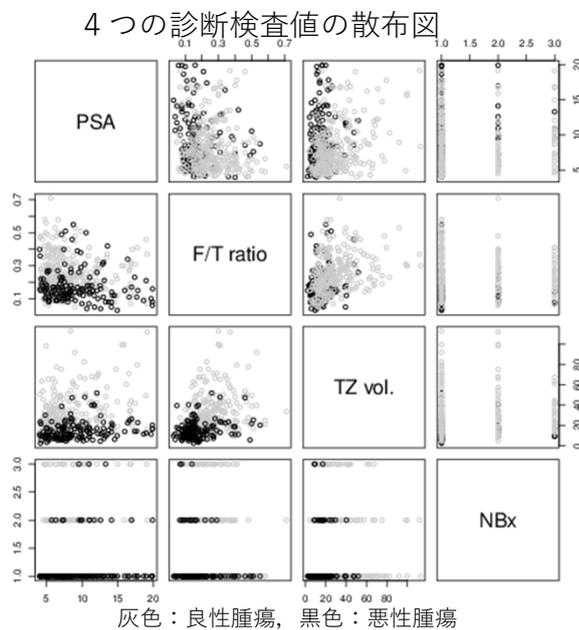
小森 理

背景

データの背後にある確率構造を捉え、高精度の予測モデルを構築したい



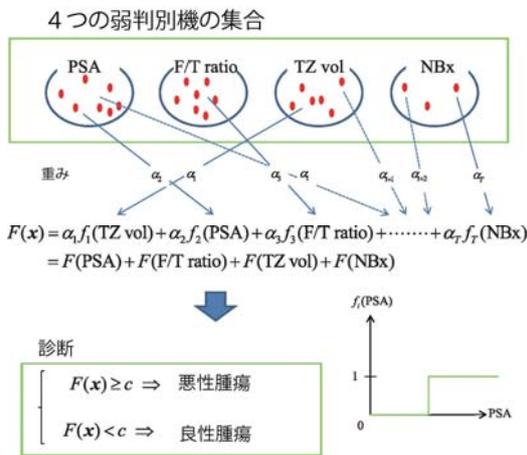
例えば病院のデータを活用すれば、患者の診断精度を向上でき、身体的精神的な負担を軽減できる。



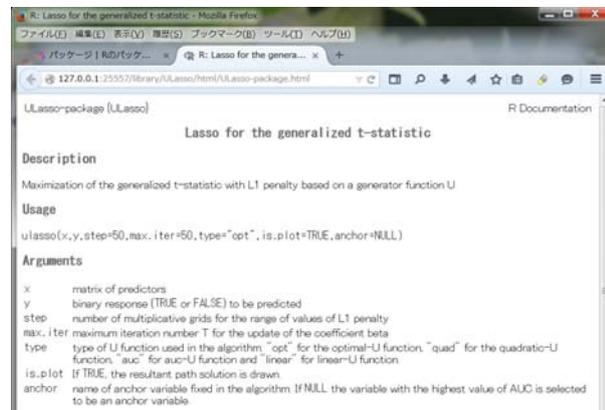
研究室の成果

さまざまな予測アルゴリズムを構築。統計解析ソフトRでpackageを作成し公開。いくつかの成果は特許の取得につながった。

予測アルゴリズムの構築



Rのパッケージのweb公開



応用例

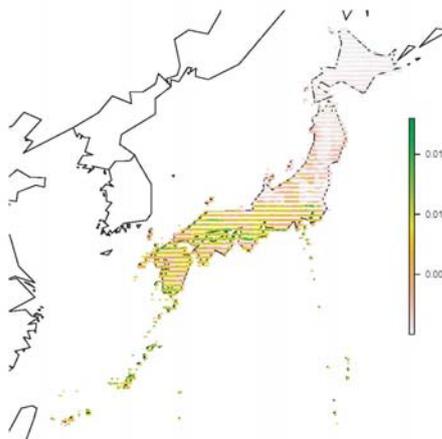
さまざまな予測モデルを構築

最適な診断閾値の推定

Table 2. Individualized PSA cut-off values for men aged ≥ 70 years old with normal DRE ($\leq 5\%$ risk of missing intermediate/high-risk cancer)

Initial biopsy	TZ volume (cc)										
% PSA	<12.0	12.0-15.9	16.0-19.9	20.0-23.9	24.0-27.9	28.0-31.9	32.0-35.9	36.0-39.9	40.0-43.9	44.0-47.9	≥ 48.0
<8	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)
8-11	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)
12-15	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)
16-19	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)
20-23	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)
24-27	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)
28-31	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)
32-35	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)
36-39	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)
≥ 40	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)	4(4-4)

日本の植物生息分布予測



世界の水産資源の予測

