



PRMLゼミ DEEPANALYTICS発表

2017年6月22日

佐藤/河井/雷/朱

役割と目指した点

- 佐藤・河井: データ加工および機械学習(SVRなど)
 - スコアの最良化を狙う
- 雷・朱: PYTHONと機械学習のチュートリアル
 - PRMLの目的...まずはPYTHON等の言語に慣れさせる

用いた手法について

重回帰
分析



リッジ
回帰



非線形手法

ランダム
フォレスト

SVR

経緯について

MRA

- 連続値予測(\Leftrightarrow 離散値・分類)のため、まずは線形回帰！

Ridge

- 過学習を防ぐために、重みの正則化項を追加！
- スコアの上昇

非線形

- SVR...初期で精度が出ないため棄却
- RMF...そこそこ期待の持てる精度のためグリッドサーチで最適化

採用データ(特徴量)について

特徴量	加工
曜日	休日かどうか、2値化
天気	悪天候(雨・雪)/屋内/それ以外 の3値
開始時刻	昼かどうか、2値化
スタジアムごとの収容人数	なし(連続値)
ホーム・アウェイチーム	あるチームが、ホーム(アウェイ)かどうか、2値化

ワンホット 変数	仙台 Home	鹿島 Home	名古屋 Home	...	仙台 Away	鹿島 Away	名古屋 Away	...
ユアテック 仙台 vs名古屋	1	0	0		0	0	1	

特徴量採用までの経緯

特徴量	除外要因
気温・湿度	相関が小さい
スタジアム名	キャパシティで代用
選手・審判	活用法が不明
住所	活用法が不明 キャパシティで代用
ゲーム開催年	3年しかない ノイズの要因と予想

結果

- ランダムフォレスト
- グリッドサーチで最良パラメタ探索
- 木の深さ15・数60

25 (-)

3,413.03020

8

2017/06/21 20:39

3413.03020

今後について

- DEEPANALYTICS: 精度の向上
 - EX.選手情報をワンホットに...など
 - すべての機械学習手法を試せたわけではない
- 雷・朱: 様々な機械学習のチュートリアルを行う
 - 後々はともに解析を行えるようにする