

J League Visitors Regression

Jリーグの観客を予想する回帰モデルを作成する。

データの概要

2012年～2014年シーズン前半戦（～7月31日）のJ1とJ2リーグの各試合の観客動員数データ
各試合のデータには、ホームチーム名、アウェイチーム名、開催スタジアム等の基本的な情報のほか、試合開催時の天候、各チームの出場メンバー、試合結果等の付加情報も含まれます。

分析の種類(回帰 or 分類など)

回帰モデルを使って予想する。
評価は RMSE（平均二乗誤差）を使って行います。

工夫すべき点

- ・特徴量の選定(膨大な特徴量をどう削減していくか)
- ・手法の選択(重回帰～ニューラルネットワークまで)



<https://deepanalytics.jp/compe/27?tab=compedetail>



Leaf Classification



葉っぱの種類を分類

データの概要

- 1584個の**画像** + **特徴量**
- 正解ラベルは99個
- 訓練データは990個 (1クラスにつき10個の訓練データ)
- テストデータは594個 (1クラスにつき6個の訓練データ)

分析の種類(回帰 or 分類など)

- 葉っぱの特徴量を用いた**分類問題**
- CNNなどのDLを用いた**画像認識**

<https://www.kaggle.com/c/leaf-classification>

工夫すべき点

- 画像の様々な**前処理**が必要
(リサイズ、グレースケール、平滑化など)
- **Augmentation**の必要性

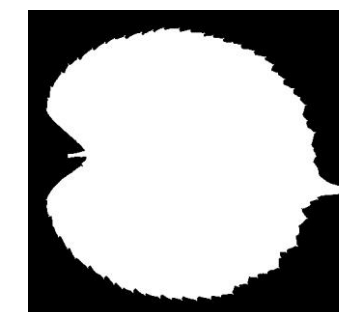
入力

出力



モデル

Acer_Opalus (1)



モデル

Tilia_Tomentosa (5)

Wholesale Customers

食品小売店舗の出店戦略を考えるためのデータ構造を浮かび上がらせる。

データの概要

小売店舗ごとの食品の売上データが、
カテゴリ別に記載されたデータ

店舗の地域や販売チャネルなどの情報

分析の種類(回帰 or 分類など)

既存の店舗をよく説明する軸を抽出し（**主成分分析**）
質の店舗をグループにまとめる（**クラスタリング**）

->教師なし学習

UCI, "Wholesale customers Data Set"

工夫すべき点

- ・データの選別(外れ値を持つレコードの削除など)
- ・分析しやすい形に**データを可視化**する。

近隣店舗の売上データ



既知のデータの構造をとらえる



とらえた構造をもとに計画を立案

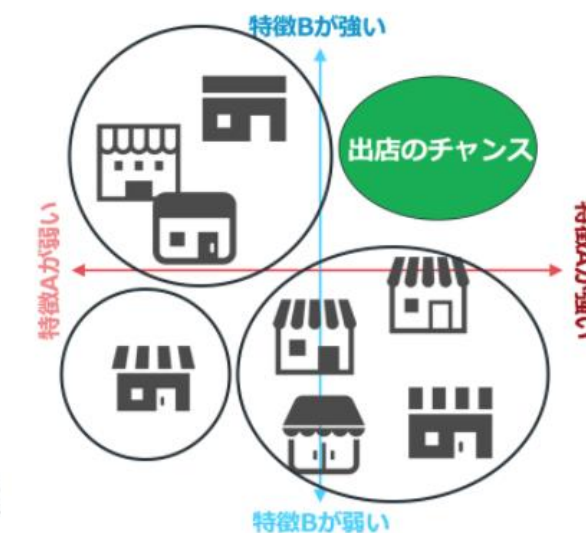


図: 店舗のポジショニングマップ



Face Gender

DLを用いて顔写真からの性別判定を行う。

データの概要

マサチューセッツ大学アマースト校が配布している
Labeled Faces in the Wildというデータセット
Webから集められた**13000枚**もの画像を集めており、
1680人の人の顔によって構成されています。

分析の種類(回帰 or 分類など)

Deep Learningの講座の5章までの知識を総動員して、
像認識の中でも特に広く実装されている**顔認識**の課題

画像データを学習させ、性別を分類する。

[University of Massachusetts Amherst, "Labeled Faces in the Wild"](#)

工夫すべき点

- ・ 画像データの前処理
- ・ ネットワークの定義
- ・ 最適化アルゴリズムの選択

