

集合知プログラミング

第8章 価格モデルの構築(2)

05T4007T 江口晃

異質な変数

- 異質な変数

- 他の変数と範囲が大きく異なる変数
- 予測するにあたって見当違いな変数

- 異質な変数の例

- これまでの変数が0~100の範囲に収まるところに、価格に影響する新しい変数にミリリットル表示のボトルサイズを導入。
- データセットに、ワインを見つけた倉庫内の通路番号が含まれる。

データセットの追加

- 異質な変数が及ぼす影響の例

- ワインセット1

- 価格に影響を与える変数: レーティング、年齢

- ワインセット2

- 価格に影響を与える変数: レーティング、年齢、ボトルサイズ

- 価格に影響のない変数: 通路番号

- K近傍法での評価比較

- ワインセット1 ⇒ 166.97339783733005

- ワインセット2 ⇒ 1427.3377833596137

- 重み付けK近傍法での評価比較

- ワインセット1 ⇒ 200.34187674254176

- ワインセット2 ⇒ 1195.0421231227463

次元のリスケール(縮尺変更)

- 値の縮尺変更
 - 不必要な変数を排除する
 - 不要な変数の計算に及ぼす影響を最小化する
 - 計算を行う前に縮尺を変える
- 値の縮尺変更の例
 - ボトルサイズの次元(軸)を1/10にスケールダウンする
 - 通路番号の変数に0を掛ける
 - K近傍法での評価 $\Rightarrow 660.9964024835578$
 - 重み付けK近傍法での評価 $\Rightarrow 852.32254222973802$

縮尺の最適化(1)

- 値の縮尺変更の問題点
 - どの変数が重要ではなく、どの変数が大きな影響を与えるのか事前に判るとは限らない。
- 縮尺の最適化
 - 考慮すべき入力変数が多数存在する場合の優れた解決方法は最適化である
 - パラメータ ⇒ ワインセット2の変数
 - パラメータの範囲 ⇒ 重み付けの範囲
 - コスト関数 ⇒ クロス評価の関数

縮尺の最適化(2)

- 縮尺の最適化の例

- 模擬アニーリングでの結果

- レーティング:11、年齢:18、通路番号:0、ボトルサイズ:6

- 遺伝アルゴリズム

- レーティング:20、年齢:18、通路番号:0、ボトルサイズ:12

- 実行例から言えること

- 通路番号が無用な変数と定めて縮尺を0とした

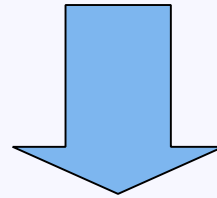
- ボトルサイズがその重要性に不釣り合いなほど大きい関数であることを見抜き、他の二つの変数の尺度を上げている。

- どの変数がどの程度重要であるか瞬時に見てとれる

不均一な分布

- これまでの前提

- データの平均または加重平均が最終価格の非常に良好な推測値になっている



- しかし、結果的に大きな影響を与えうるが計測されていない変数が存在するような状況もある。

- 例: ワインを酒屋で購入する客と、ディスカウントストアで購入して40パーセントの割引が受けられる客がいる。

- 推論結果は両グループのアイテム平均、20パーセントの割引になる。

確率密度の推測

- 確率密度の推測

- 指定範囲内にある近傍群の重みの合計を、すべての近傍群の重みの合計で割る。これによって、あるアイテムが特定の価格帯に落ちる確率がわかる。

例: ワインのレーティングが99で20年のとき、その価格が40~80ドルの確率が50パーセント、80~100ドルの確率が50パーセントである。

実際にプログラムで実行した結果

40~80ドルの確率 ⇒ 0.62305988451497296

80~120ドルの確率 ⇒ 0.37694011548502687

120~1000ドルの確率 ⇒ 0.0

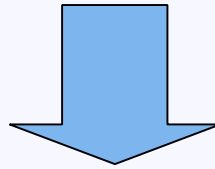
30~120ドルの確率 ⇒ 1.0

- 範囲を小さく分割してやれば、アイテムが実際に密集している範囲を定めることができる。

確率のグラフ化(1)

- 確率密度の推測の問題点

- データが取る構造のクリアな像が得られるまで、自分で範囲を推測して入力する作業を繰り返す必要がある。



- 確率密度をグラフとして描けば、考慮に入れるべき範囲を推測しなくてもよくなる。

- 累積確率のグラフ
 - さまざまな価格点の実際の確率をグラフ化する

確率のグラフ化(2)

- 累積確率のグラフ

- グラフの値は0から始まり、価格を持つアイテムに出会うたびに増大する。最大の価格に達したときに1となる。

確率のグラフ化(3)

- さまざまな価格点の実際の確率をグラフ化する
 - 確率をそのままグラフ化するとほとんどの場所で0になるため、一定の範囲の確率をグラフ化することで見やすくする。