

データサイエンス概論

第5回 第1講

事前予習

インターネットで、以下のようなキーワードで検索を行い大まかな内容に目を通しておく。

- 判別分析
- ロジスティック回帰

第5回で取り上げる分析手法は、どれもよく使われるものであり、この講義に続くデータサイエンス関係の講義でもより詳しく取り上げられる。本講義では概要が掴めればよい。

注：

スライド：1-10

分散を、 N で割るのか $N-1$ で割るのか、という点については長い議論がある。ここでは $N-1$ で割るという多くの教科書で採用されている方にしておく。

スライド：1-11

訂正：LDS(Liner --> LDA(Linear Discriminant Analysis) のための関数
(ライブラリとしては、MASSというライブラリに入っている。5-LDA-iris.Rmdスクリプトを参照)

データサイエンス概論

第5回 第2講

事前予習

インターネットで、以下のようなキーワードで検索を行い、大まかな内容に目を通しておく。

- クラスタ分析
- ユークリッド距離
- 正規化

データサイエンス概論

第5回3講

事前予習

インターネットで、以下のようなキーワードで検索を行い大まかな内容に目を通しておく。

- 決定木
- エントロピー
- 情報利得

検索してみると、決定木のノードの呼び方などにバリエーションがあることたわかる。

決定木については本講義以降のデータサイエンス系の講義でも取り上げられる予定であるが、決定木の学習アルゴリズムの詳細に興味のある人は、以下のような項目を検索してみることに。

- [ID3](#) (Iterative Dichotomiser 3)
- [C4.5](#)
- [CART](#) (Classification and Regression Trees)
- [CHAID](#) (Chi-squared Automatic Interaction Detection)

データサイエンス概論

第5回 4講

事前予習

インターネットで、以下のようなキーワードで検索を行い大まかな内容に目を通しておく

- 顧客分類
- 銀行 融資判断 (貸付判断)