

統計工学実験 (Statistical Engineering)

w.namamoto

はじめに

この実験では3週間に渡って、ひとつのデータ分析(データ解析、データサイエンス、データマイニング)の問題に取り組む。3週間で凡そ次のような段階を辿ることを想定している。

1. データの理解 (現状把握)
2. データの解析 (データ解析、データマイニング、データサイエンスの第1段階)
3. データからの (役立つ) 知識の抽出 (データ解析、データマイニング、データサイエンスの第2段階)

データ解析の手法は、ただ闇雲に用いただけで必要な知識がデータから抽出できるというものではない。様々な手法の、原理、仮定(前提)、その手法が用いるモデルの構造とパラメータの意味を知り、それぞれの長所と短所を理解した上で、解析するデータに適切な手法を用いるべきである。そのためには様々な手法を適用する前にまず、データそれ自体の理解が不可欠である。今回はこれを目的として、解析の第一段階としてのデータの理解に努めてもらった。

今週はリンク先の同志社大学の金(じん)先生が公開してくださっているコンテンツを参考に、判別分析と決定木(樹木モデル、分類木とも)を用いて、今回のデータを解析して貰う。解析の目的は「定期預金を契約する人と契約しない人の間に、どのような差異があるか」を調べることである。

<http://bit.ly/mselab2013stat> (統計工学実験)

<http://bit.ly/mselab2013> (経営情報学実験全体の情報ページ)