

データサイエンス概論

始めに

「データサイエンス」は、新しい領域であると同時に、データ分析の対象となる領域も、分析に用いられる手法もいくつもの学問領域をまたがるものが多い。その全てを概観することはできないので、一定の方針を決めて見ていくことにしたい。

1、本講義は1年次の1学期に学ぶ学生もいることから、主な対象を本学入学以前に統計、プログラミング、コンピュータサイエンス等にほとんど触れたことのない人とし、これから出会う概念や考え方に触れながら、データサイエンスの全体像を見ていくことを目的とする。

2、基礎の部分に比較的時間をかけた後、応用例を見ていく。基礎の部分とは言っても、範囲が広く、これまでとは異なった考え方を要求されることも予想される。特に全く予備知識のない部分については、指示された予習を行なっておくこと。

3、この講義でデータサイエンスでよく用いられる概念や分析手法の輪郭を理解し、続いて、人工知能概論、社会データ分析、計量テキスト分析、データサイエンス演習、その他プログラミング科目などで詳細について学んでいくこととなる。

4、授業の説明に用いているデータの分析のサンプルのほとんどは、Python言語かR言語を用いている。用いたスクリプト（≒プログラム）は、参考のためクリックすれば自分で動かしてみることのできるNotebook形式で提供する。これはあくまで参考のために提供するもの

で、小テストや単位認定試験の対象にはならない。また、インストールされている機械やパッケージのバージョンなどの変更によって動かないこともあるので、スクリプトが動かないという問題へのサポートはできない。

以上のような方針のため、すでにコンピュータ関係の仕事についているというような人には退屈な部分もあると思われるが、そのような人はぜひ他の受講者のサポートに回って欲しい。

事前予習について

データサイエンスの入門書は既に色々出ているが、上に述べたようなこの講義の目標に沿った一冊のテキストを探すのは難しい。さらに、オンラインの大学である本学の図書館の特性、そして著作権の問題から、インターネットで関連のあることを検索し、背景として大まかなイメージを掴んでおいてもらうこととしたい。検索して出たきたものを全て理解する必要はないので、一般にどのような話題が出てくるかを見ておいてもらえれば十分である。

また、本講義はデータサイエンス概論の全体像を展望するという趣旨に加え、15分ずつの授業ということもあり、個々の概念の説明に十分な時間を割くことが難しい。いくつかの授業については補足の情報やリンクを「補足情報」として追加しておくこととする。

*複数の検索キーワードの間に、「データサイエンス 歴史」のようにスペースのあるものは、スペースで区切られたキーワードを全て含むページを検索するAND検索を行うことを意図している。通常はスペースが入っていればAND検索と解釈される。

データサイエンス概論

第1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。

- データサイエンスとは
- ビッグデータ
- データサイエンス 入門

授業の始まる前に、現在データサイエンスという言葉がいかに広範囲に用いられているか、データサイエンス入門としてどのような内容が一般に触れられているかが分かれば十分である。また、入門書やネットの解説にどのようなものがあるか、一般的なデータサイエンスの入門的な解説の雰囲気を見ておくこと。内容を理解する必要はない。

注：

スライド1-3:

訂正：「定義」 → 「定義と歴史」

データサイエンス概論

第1回 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。

- データサイエンス 応用 領域
- もしくは
- データサイエンス 活用 領域

どちらのキーワードでも良いが、データサイエンスが一回の授業でカバーできないほどの多くの領域で活用されていることがわかれば十分である。

データサイエンス概論

第1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。

- ・ データサイエンス 学際領域

データサイエンスが、いろいろな分野に跨る、いわゆる学際領域である、という点がいろいろなところで指摘されている。どのような領域が関わっていると主張されていくか、大まかに見ておくこと。

データサイエンス概論

第1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。

- ・ データサイエンス シラバス

日本の他の大学でどのような内容の授業が行われているか、いくつか覗いてみることに。

ついでに、

- ・ data science syllabus

とすれば、世界の大学でのシラバスを検索してみることができる。あまりにヒットの数が多くて圧倒されるかもしれないが、世界中で、また多様な学部でデータサイエンスの授業が開講されていることがわかる。

サンプルスクリプトについて

第1回第3講のスクリプトをRStudioのnotebook形式で保存したものをダウンロードできるようにしてある。

DS1-3_script1.Rmd

DS1-3_script2.Rmd

これを試してみるには、RStudioというプログラムをインストールしてファイルを開く必要がある。インターネットにインストールの方法の説明はたくさん出ているが、例えば、

https://qiita.com/daifuku_mochi2/items/ad0b398e6affd0688c97

を参照。