

データサイエンス授業担当報告

～ 文系? 1年生にデータサイエンス教えてみた ～

筑波大学 システム情報系
学術情報メディアセンター 亀山啓輔

筑波大学 共通科目「情報」 データサイエンス カリキュラム

対象 全学群(学部)1年生 (必修)

概要

- データサイエンスの基礎的概念を理解し、コンピュータを利用した基礎的なデータ分析技術を学ぶ.
- データの収集, データの管理, データの可視化, データの分析を通じてデータの理解と活用の手法を実践的に修得する.
- 先端的なデータサイエンスの事例に触れ, 社会におけるデータの具体的な活用について理解する.

授業形態

- 講義(座学+ビデオ講義(動機付け, ビッグデータ, AIと機械学習))
+ PC演習(Excel初歩, データ収集, 可視化, データの種類, 統計量)

筑波大学 共通科目「情報」 データサイエンス カリキュラム

各週の内容 (10週. 毎週講義1コマ＋演習1コマ. 2単位)

1. イントロダクション (動機付けビデオ視聴)
- 2～4. データの種類と収集
(データの種類, 収集と調査, 個人情報と研究倫理, 処理の手順, 誤差, データ管理, アンケート, 関係DB)
5. 高度なデータ管理と活用 (ビッグデータ & IoT ビデオ視聴, 中間試)
6. データの可視化 (クロス集計表, グラフ, 可視化の作法)
- 7～8. 質的／量的データと統計量 (ヒストグラム, 代表的な統計量)
9. 様々なデータ分析 (時系列解析, 移動平均, 指数化, 分析の設計)
10. 高度なデータ管理と活用 (人工知能と機械学習 ビデオ, 期末試験)

共通教材

パワーポイント資料, 演習用エクセルファイル, ビデオ教材

(ビデオ教材はOCWとして公開 <https://ocw.tsukuba.ac.jp>)

筑波大学 社会・国際学群 国際総合学類

学類教育

- 国際的な諸問題の解決・分析を行う人材育成を目指した文理融合教育.
- 国際政治・国際法, 経済, 文化・社会開発, 情報・環境の教員

学生の性質, 気質

- 入学時(入試)は大半が文系だが理系科目にも興味はある
- 国際的なキャリアを志向.
- コミュニケーション・語学能力高い. 卒業までに半数が短期留学.
- 情報収集能力は高い. 対人, 対外国語ハードル低い
- 複雑な対象の全体像を(ざっくり)把握する能力にたける. 構成部品の動作原理や仕組みにはあまり興味なし.
- 元気. 授業中さわがしい.

「データサイエンス」への関心は高い

初回宿題(動機付けビデオ視聴後)

データサイエンスの知識やスキルは、あなたが国際総合学類で学び、自らの専門としたいことにどのように関係しているか(関係していそうか)を記しなさい。

解答抜粋

• なんでデータサイエンス？(少数派)

「… 国際総合学類で学べることに特に関係していることなどはいまいち思いつかなかった。」

• 目指す専門性におけるツールとして(多数派)

「**国際政治学**が研究対象とする様々な事象に関してより客観的で信ぴょう性の高い考察をデータに基づいて導く上で重要だと考える。」

「…**教育開発**を専門的に学びたい。教育開発はその効果を探るうえで統計をとることが欠かせない。」

「…**農業経済**では、いつ収穫して市場に出すかなどのデータや、農家の声などの情報を蓄積して考えることが必要だと思う。それらの情報を扱いやすくしたり、必要な情報を容易に取り出したりする必要があるので、データサイエンスの知識があれば少しはやりやすくなると思う。」⁵

「データサイエンス」への関心は高い

解答抜粋(続)

- 広く情報を取り入れ、分析し、発信するプロトコルとして

「…世界で起きている様々な課題や問題の解決には、まずはその問題の背景や要因を知る必要があるが、その段階でデータサイエンスのスキルを使用し、**合理的に考える**ことができるのではないか」

「…社会科学とは社会という現実にもとづいた研究が必要であり、**現実**は**数値として扱う**のがもっともわかりやすい。」

「…データサイエンスという手段を駆使して正確な世界環境の現状を把握することが必要であると考え、**現状の正しい把握**が出来ていないと解決のための行動を誤ってしまう。」

「… 私が例えば日米の関係を研究するとしたら、… 過去の多量のデータを集めて分析する、… そうすれば、データを視野に入れずに論じるのとは比べて確実により詳しい**根拠になる**ことができると思う。」

学生との雑談から

- 「データアナリスト」はカッコよさそう
- データサイエンスわかっていると就職に有利そう

講義雑感

全体

- 座学で学んだ内容をすぐにPC演習で確認するサイクルは有効
- 専門性に関連する知識やスキルに関する内容は好評.
(国総生では, アンケートの取り方, データの相関解析)
- 高校での情報, 数学の学習内容にばらつきあり. 授業はゆっくり進め, 既学習者の演習先行は自由にした.

履修後アンケートから

16. 中学・高校の授業におけるデータ解析ソフトの学習内容と今回の講義内容はどの程度重複していましたか。

回答	人数
全く重複していない。 / No duplication at all.	3
ほとんどは初めて学ぶ内容だった。 / Slightly duplicated, but mostly new.	18
半分程度は重複していた。 / 50 %% duplicated.	8
多くは重複していた。 / Mostly duplicated.	0
全て重複していた。 / 100%% duplicated.	0

講義雑感

エクセル演習

- リテラシ科目として表計算, グラフ可視化などの機能を一通り教えることは重要(興味はまちまち)
- 筆算や電卓ではとても扱えない量のデータで演習を行わせることは有効(自分の能力を拡張するツールとして認識できる)
- ネットからデータを取得して行う課題は好評. ごく基本的な課題にもそうした要素があってよい. 海外のデータも活用したい.

履修後アンケートから 履修前と履修後のデータ可視化・分析の習熟度に関する自己評価

	履修前	履修後
1. データ分析ソフトを用いた事がない。	8	0
2. データ分析ソフトを用いて、グラフ機能を使って、データの属性、変数の数に合わせたグラフが作成できる。	17	10
3. 2に加えて、データを代表値を使って表現したり、データのばらつき方を議論したりできる。	2	6
4. 2.3.のスキルを利用して、与えられたデータの特徴や傾向を説明することができる。	2	13

講義雑感

ビデオ講義

- 学生の興味や志向する専門分野とデータサイエンスを結びつける動機付けは重要.
- トピックを絞り、解決すべき問題を明確にしたものを多様な興味にこたえられるよう多数準備するとよい.
(いくつかは学生の興味を引くものがあるように)
- 発展的内容(ビッグデータ, 人工知能)は, 理解できたものの, 「自分がすぐに使える感」は薄かったよう.

講義・演習環境

- 話し合い, 教え合いながら演習を進められる環境が理解の助けにもなり, 望ましい.
 - 新入生ゆえ, なおさらそうした機会が重要.
 - コロナ禍の下, 遠隔授業で実施するのはつらい.
- 巡回するTAは質問しやすく有用.

高度な内容

ビデオ教材 馬場雪乃先生 「人工知能と機械学習」

YouTube JP

検索

ログイン

機械学習のしくみ：訓練データ

訓練データは数値で表現されないといけない

訓練データの例

	入力		出力
	大きさ	でこぼこ度	
腫瘍1	0.252	0.014	良性
腫瘍2	0.327	0.340	良性
腫瘍3	1.000	0.371	悪性
腫瘍4	0.223	0.369	悪性
...	...	...	...

図示

● 良性の腫瘍
● 悪性の腫瘍

でこぼこ度

大きさ

2'

32:26 / 57:35 • 機械学習のしくみ：訓練データ

高度な内容

期末試験問題から（正答率約8割）

(Q10) 2つの量 (x, y) を手掛かりに, (ビデオ講義で示された良性腫瘍と悪性腫瘍のような) 2つのデータ群を識別する規則を, モデル

$$z = wx - y + b$$

のパラメータ w と b の値を変化させて決定したい. パラメータを決定したのちに, この規則を新たな観測 (x, y) に適用する際には, $z < 0$ なら群 1 に, $z \geq 0$ なら群 2 に属するものと判定することにする. 今, 図 3 に各群 8 組の実例が群 1, 群 2 それぞれについて, 「○」と「▲」で示されている.

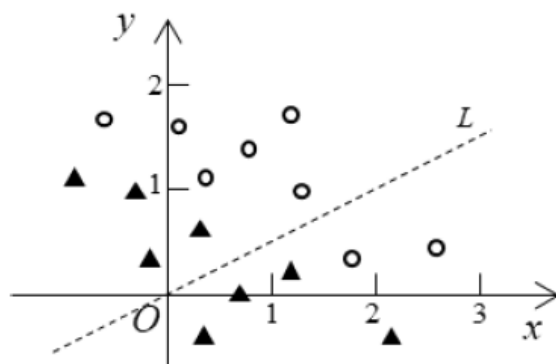


図 3 2 群のデータ分類規則を学習するための実例各 8 組. 群 1: 「○」, 群 2: 「▲」

- (a) 図中の点線 L は, $w = 1/2$, $b = 0$ としたときの $z = 0$ となる点の集合である. 損失関数を「与えられた実例のうち, 誤って分類してしまうものの数」と定義するとき, 点線 L の規則を用いた場合の損失を求めよ.
- (b) 図中の実例の分布を手掛かりに, 損失が最小となるパラメータ w と b の値を求めよ.

まとめ，課題

- 情報リテラシ教育の一つとして，データの基本的な扱いに焦点をあてた科目設定は，理系・文系を問わず有益.
- 2年次以降の多くの専門科目へのハードルを下げた（のでは？）
- 多くの分野の調査や研究においてデータに基づく客観性が求められており，多量のデータに基づく観測，分析，情報発信のプロトコルを必須のものとして示し，その基本的手段を与えることは重要.
- 発展科目を準備してそれぞれの専門分野への円滑な接続を図る必要あり.