Data Mining Theory

DM-01-intro: Introduction

Class 1: Shugo NAKAMURA

　@Shugo NAKAMURA　shugo.nakamura@iniad.org

Class 2: Takashi SANO

　@sano　takashi.sano@iniad.org

● Business Innovation Course / Data Science Field

● Please let us know via slack if the streaming of video and audio are not stable.

Preparation

- Do you check your iniad email and lecture slack at least once or twice a day?
  - You can set up an email notification for any slack posts that are left unread

- Have you register to the channel of this lecture on INIAD-lectures slack?

- Please open a window for INIAD-lectures slack besides Meet window.
  - Write to slack channel like to speak in the classroom at any time during the lecture.

- Have you open the folder for this lecture on Google drive?

- Since you need to view many windows at the same time, make sure you use your tablet or smartphone as well.
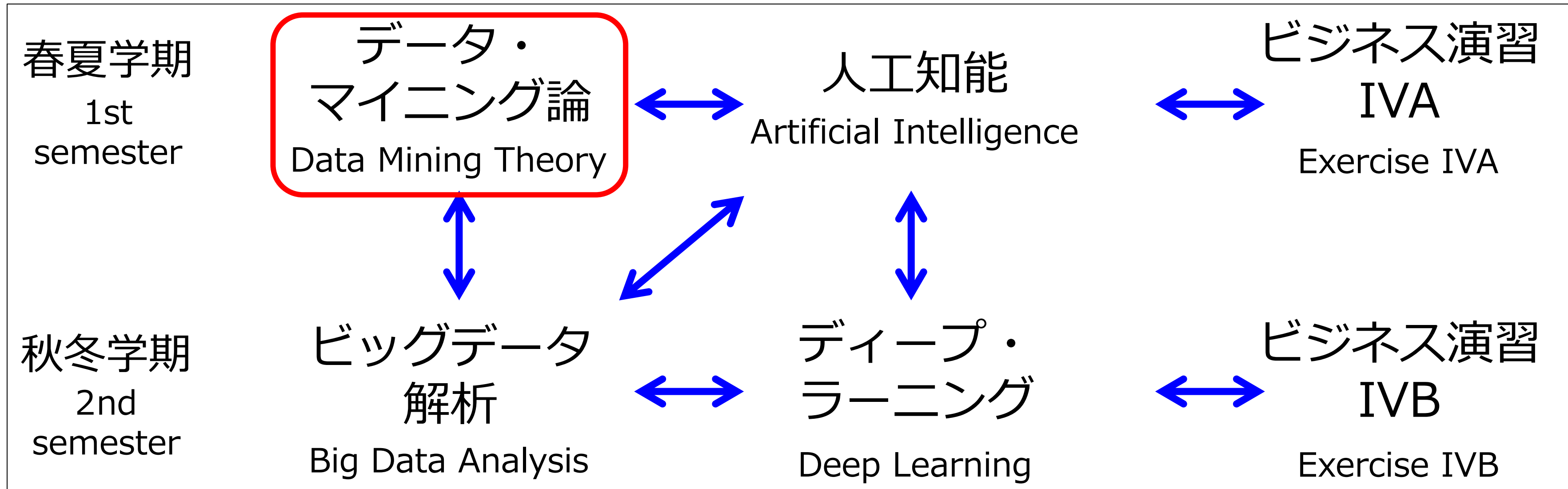
ビジネスコース データ・サイエンス系科目
Subjects in Data Science Field / Business Innovation Course

2年次 2nd grade	統計とデータ分析I,II Statistics and Data Analysis I,II	ビジネス演習IIA,B Exercise IIA,B

3年次
3rd grade

春夏学期
1st
semester

データ・マイニング論
Data Mining Theory ⟷ 人工知能
Artificial Intelligence ⟷ ビジネス演習
IVA
Exercise IVA

秋冬学期
2nd
semester

ビッグデータ
解析 ⟷ ディープ・
ラーニング ⟷ ビジネス演習
IVB

Big Data Analysis　　　Deep Learning　　　Exercise IVB

Themes (May be changed in the future)

Visualization
Review of Statistical test
Bayesian classifier
Dimensionality reduction
Factor analysis
Recommendation system
Association analysis
Optimization

到達目標
Learning objective

典型的なデータ解析法はサンプルプログラムが入手できることが
多いことを踏まえて…
Considering that sample programs are usually available …

● 各方法の動作原理を理解し、自分の問題にどの方法が適しているかを**判断できる** Understand the principle of each method and become to be able to determine the suitable method for a given problem

● サンプルノートブックの内容を把握し、**方法とPythonプログラムの対応関係**を理解する Understand each line of notebook and correspondence between the method and the Python code

● **自分のデ**ータに対してノートブックを適用できる・**自分で**プログラムの間違いをみつけて修正できる Apply notebook to your own data / correct your code

● ノートブックを変更して**機能変更・拡張**ができる Modify / enhance notebook

受講要件

Pre-requisites

ビジネスコース以外の皆さんへ

For students other than business course

- 要素技術が相互に関連しあうことが多いので、他のデータサイエンス系科目を受講しない場合でも、MOOCsなどを活用して**自学自習することを強く推奨**する

  It is **strongly recommended to do self-study** using MOOCs etc. for other data science subjects, because elemental technologies are often related to each other

- とくに、3年次前期「データ・マイニング論」「人工知能」の最初の数回は、**共通の重要な要素技術の復習**を含めた学習を予定しているので、必ず理解しておくこと

  In particular, the first few lectures of "Data mining theory" and "Artificial intelligence" are to learn common important techniques and concepts, so be sure to understand

成績評価 Grading

● ビジネスコース学生は必修

Compulsory subject for Business course students

● 試験、理解度確認小テスト、課題提出、講義への積極的な取り組みな どを総合的に評価する予定

Evaluation is based on the exams, comprehension level check quiz, assignments, positive attitude to lectures, …

● 単位取得に最低限必要な出席回数は、総回数の3/4以上を目安とする

The minimum number of attendance required to get the credit is, in principle, 3/4 of the total.

Flow of the study of each lecture

Prepare for the contents of MOOCs for **DM-(i)** and do its assignment (submission deadline: two days before the next lecture)

Report the items for additional explanation

DM-(i) lecture: quiz for **DM-(i-1)**, additional explanation for the contents and explanation for the assignment of **DM-(i)**

Revise and resubmit the assignment of **DM-(i)** (2 days before the next lecture)

Understand the contents of MOOCs for **DM-(i+1)** and do its assignment (submission deadline: two days before the next lecture)

Report the items for additional explanation

DM-(i+1) lecture: quiz for **DM-(i)**, additional explanation for the contents and explanation for the assignment of **DM-(i+1)**

⋮

- Please prepare for the class using MOOCs, submit assignments as much as possible, and report any items that are difficult to understand. The lecture will focus on the key points based on them.

Since DM-0102 is the first lecture ...

Let's see MOOCs DM-0102 (two lectures are merged)

- After this lecture (1st half of DM-0102), do review this lecture and also preparation for the 2nd half of DM-0102, and submit the assignments for DM-0102 as many as possible (deadline: Apr 14th Wed 23:59)
  - If there are parts of the contents that you can't understand well, report them from "Report on difficult-to-understand parts".

- Next lecture (Apr 16th Fri): explanations of 2nd half of DM-0102 and all of the assignments in DM-0102 (you should do self-scoring of assignments)

- Revise and resubmit the assignments of DM-0102 if needed, and also do preparation for DM-03 + submit the assignments as many as possible (deadline: Apr 21st Wed 23:59)

- For DM-0102, submission or re-submission of the assignments is regarded as attendance.

- The comprehension quiz will be start from DM-03.

After DM-03...

- Attendance will be counted by submitting both the quiz at the beginning of the lecture and the miniquiz during the lecture. Failure to submit both will be treated as an absence.

- Submission of assignments (not the re-submission) will be graded positively.

- From DM-04, there will be a "self-chosen project (free assignment)" in each class. Submission of these will be graded positively.

About the exams

(This may change in the future.)

- There will be three exams: mid-term (first half), end-term 1 (second half), and end-term 2 (whole areas). At least end-term 2 will be face-to-face.

- Since the end-term 2 is already like a "re-test", there will be no further exam (students who retake this class should be especially careful!).

- Many submissions of the assignments cannot recover the very bad scores of exams.

- If the status of the submissions is very poor, you may not be able to get credit even if the scores of exams are at the passing level.

Policy of face-to-face

- Retake students: face-to-face each time

- Other students: half of the students will be face-to-face.

- Details will be announced via slack.

その他の大切なこと　Other important things

● ハードウェアのトラブルは早いうちに解決すること。講義中に、正しい操作を行っているのにPCが意図どおりに動作しないことは、履修の大きなハンディとなる。　Use reliable hardware!

● タイピングが遅い人は、タッチタイピング（キーボードを見ないで入力）の練習もしておくとよいだろう。

　　■ 例: https://www.e-typing.ne.jp/english/　Training of touch typing

● データ解析の方法は、どんなデータにも適用できることが多い。しっかり身につければ、あらゆる分野で活用することができるので、積極的に取り組み、また身近な応用を考えてみてほしい。

The way of data analysis can be applied to any data, so let's consider your own applications