

AI / デジタル社会を担う人材育成と教育体系
～新課程における統計・データサイエンス教育の実践と高大社接続・産学連携授業～

東北大学データ駆動科学・AI教育研究センター での取り組み

東北大学の数理・データ科学教育 (入門～応用基礎)

東北大学データ駆動科学・AI教育研究センター
データ科学教育研究部門
磯辺 秀司

AIMDの基礎

「情報とデータの基礎」(所属学部に関係なく全学生必修)



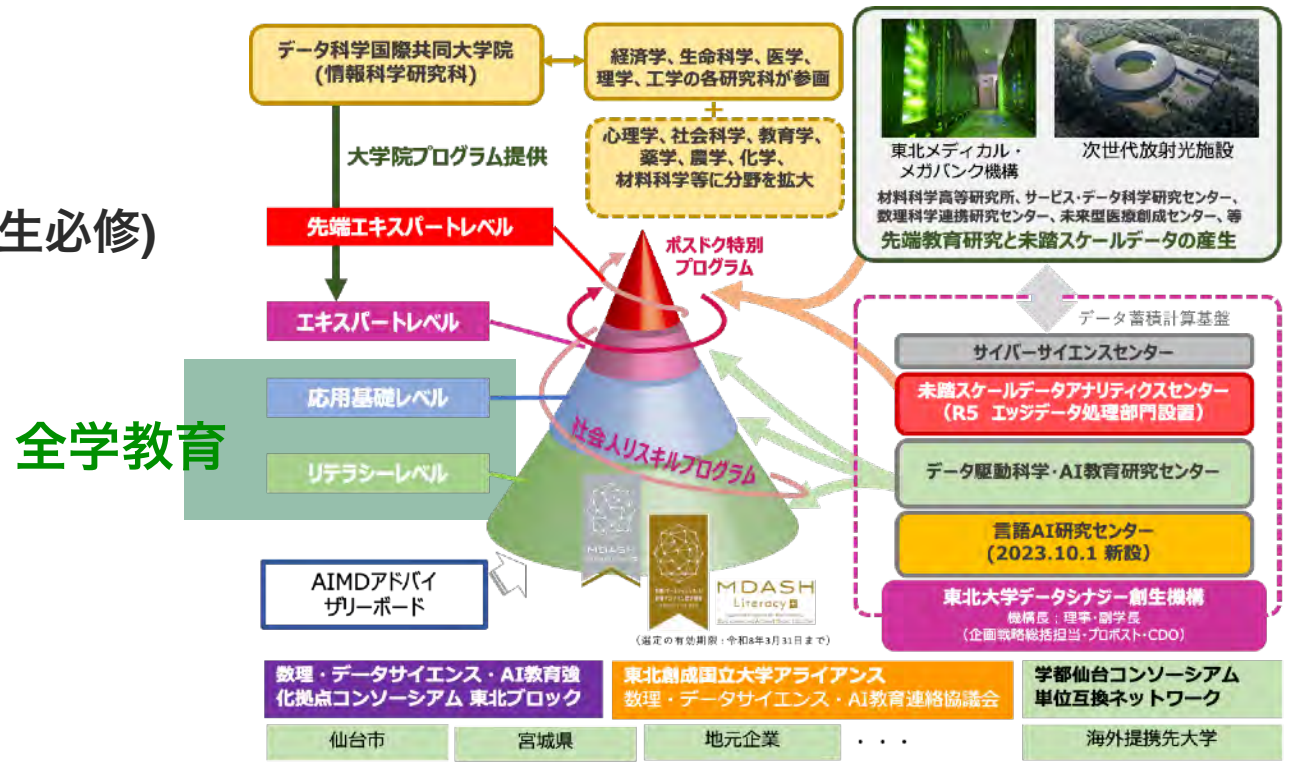
AIMDの応用基礎 挑創カレッジ CDS

意欲的な学生へ一段深い学びを提供

応用基礎レベルモデルカリキュラムのオプション内容までを含むやや高度な内容で実施

- 情報群・統計群・数学群の科目から所定の必修科目・選択科目の単位を取得することで修了認定
- 必修科目には、基本的な機械学習アルゴリズムの理解や Python を用いた機械学習実習に関する科目が含まれる。

新入生から大学院生に至るまで一貫した教育プログラムを実施
eLearning教材 (動画, OCW, MOOC, etc) などを活用した自修環境も提供



2014.11 「情報基礎」 第4版 策定

- 高等学校で教科「情報」が再編されたことに対応する改訂

2016.4 「情報基礎」 第4版 授業開始

情報活用・情報処理・情報倫理の3本柱
+ 計算機科学の思考法を活用した問題解決

- 理系・文系で (ほぼ) 共通の教育内容

2019.4 挑創カレッジ (CDS) 開始

- 数理・データサイエンス教育全国展開事業協力校
- データ駆動科学・AI教育研究センター設置 (10月～)

2020.4 「情報基礎」 の強化

- 内容を一部改変しデータ科学・AIに関する教育内容を導入。新入生全員が履修するデータ科学・AI関連基礎科目として開始

2021.8 MDASH Literacy+ 認定 (5年間)

2022.4 「情報とデータの基礎」 授業開始
(現行)

- 「情報基礎」 を発展させた形で新カリキュラムを開始
- 全学生必修

現行科目は2022年4月から**全学生必修科目**として実施

前身科目「情報基礎」の基幹を継承しつつ, AIMD基礎科目としての内容をカバー

- ・ リテラシーレベルモデルカリキュラムに対応

Social Skill

- 情報化社会を生きる市民としての責任・意識づけ, 法令の基礎知識
- サイバーセキュリティの基礎知識と意識づけ

Science Skill (1)

- 計算機科学的思考法による問題解決 (問題の抽象化, アルゴリズムの構築)
- 計算機の計算能力を活用した情報処理・問題解決 (プログラミング)

Science Skill (2)

- 実社会におけるAI・データ科学技術の活用事例 (動機づけ)
- 基本的な統計量に関する基礎知識
- 実データを用いた初等的な統計処理実習 (表計算, Python などの活用)

情報倫理

情報化社会で適切に生きていくための倫理

情報の科学

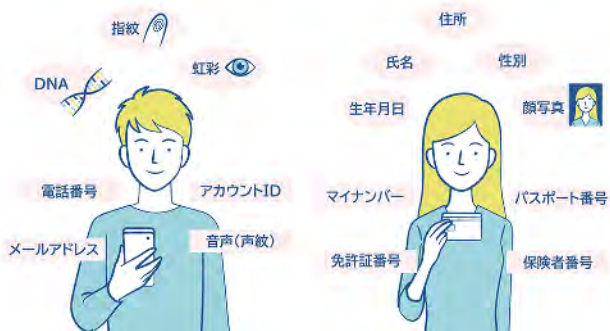
計算機の能力を活用した問題解決

データの科学

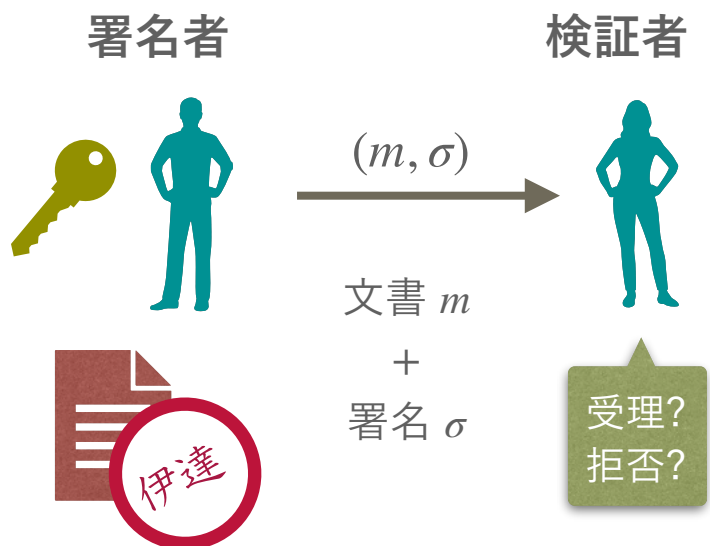
データ科学への動機づけとデータ処理への入門

Social Skill

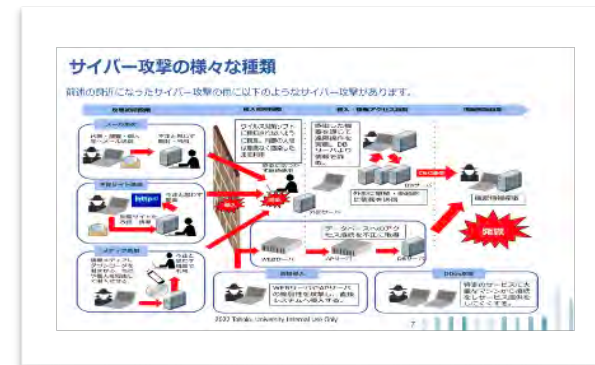
- 情報化社会を生きる市民としての責任・意識づけ, 法令の基礎知識
- サイバーセキュリティの基礎知識と意識づけ



個人情報保護, 知的財産権の仕組みなど, 情報技術周辺法令の基礎



情報セキュリティの基礎 (暗号と認証, 詐欺などの脅威への対策など)



企業との共同制作によるオンライン自習用教材を提供



MOOC講座 (東北大学サイエンスシリーズ)

Science Skill (1)

- 計算機科学的思考法による問題解決 (問題の抽象化, アルゴリズムの構築)
- 計算機の計算能力を活用した情報処理・問題解決 (プログラミング)

Computational Thinking

- 抽象化 (Abstraction) と自動化 (Automation) を特徴とする思考スタイル
- 数学的思考と完全に同じではないにせよ、相当に重なっている



センター教員執筆の教科書
(参考書として授業利用)

何人かの参加者が集まって1個のホールケーキを分けるとき、各人が満足するような分け方は？



抽象化

数学的な定式化

何人かの参加者が集まって1個のホールケーキを分けるとき、各人が満足するような分け方は？



このようにカットしたとき → ケーキ断片 I, II, III に対する各人の「満足度」(満足度はケーキ全体で1に正規化)

解法的设计と実装

処理の自動化

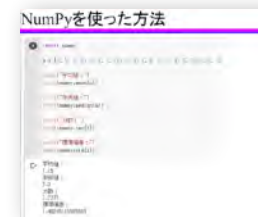
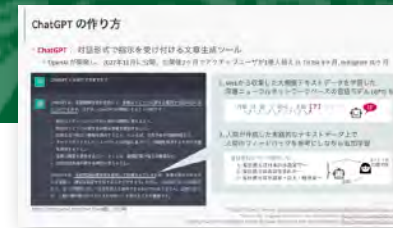
```
1 000001 cod_01_01_01
2
3 Dim cmd As SublimeText.Command = D:\Common-Tools\TextCommunity\Sublime, 000
4 cmd.Parameters.AdditiveValue("001", cod_01)
5
6 Dim dStart As DataTable = D:\Common-Tools\ExcelData\Start\PreProcess, 000
7 If dStart.Rows.Count > 0 Then
8   If Not dStart.Rows(0).IsDBNull("cod_area") Then
9     l.Add(CType(dStart.Rows(0).Item("cod_area"), String))
10  End If
11 End If
12
13 If l.Count = 0 And dStart.Rows.Count > 0 Then
14   Dim cmd_grmmer As String = ""
15   Dim cmd_grmmer As String = ""
16   If Not dStart.Rows(0).IsDBNull("cod_grmmer") Then
17     cmd_grmmer = CType(dStart.Rows(0).Item("cod_grmmer"), String)
18   End If
19   If Not dStart.Rows(0).IsDBNull("cod_grmmer") Then
20     cmd_grmmer = CType(dStart.Rows(0).Item("cod_grmmer"), String)
21   End If
22 End If
```

※ 必ずしも計算機による処理を意図しない
(実装技術はオプション)

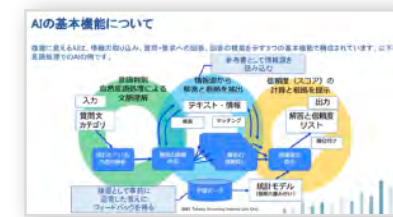
Science Skill (2)

- 実社会におけるAI・データ科学技術の活用事例 (動機づけ)
- 基本的な統計量に関する基礎知識
- 実データを用いた初等的な統計処理実習 (表計算, Python 等の活用)

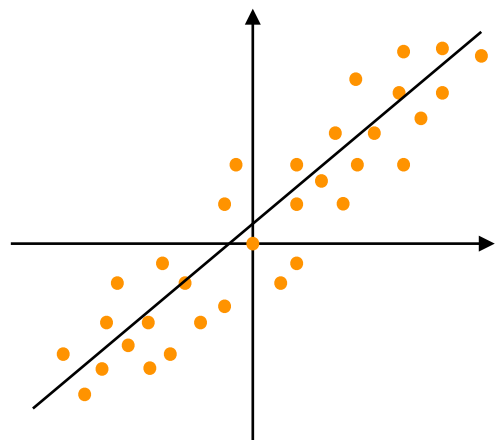
補助教材
「対話型AIの活用」



動画教材
「Pythonによる統計学入門」



動画教材
「人工知能(AI)基礎講座」



実データを用いた実習



```
1 listRaw=[]
2 listRawy=[]
3
4 for i in range(0,len(listRaw),1):
5     listRawx.append(listRaw[i][0])
6     listRawy.append(listRaw[i][1])
7
8 plt.figure(figsize=(40,10),dpi=50)
9 plt.plot(listRawx,listRawy)
10 plt.plot(listMaxv,listMavy)
11
12 plt.show()
```

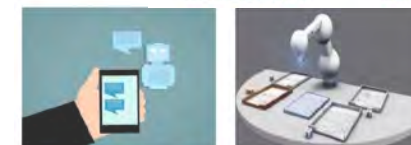
日経平均株価の移動平均を算出して視覚化。値動きの大雑把なトレンドを観察

- 平均, 分散, 共分散, 相関係数の定義と意味
- 「平均は常に代表値か?」「相関は因果関係を意味するか?」など, 統計量の読み方

- Python (または表計算) を利用して基本統計量を計算し, そこから何が読み取れるかを考えさせる



家電製品にAI 日常生活を便利に
教育にAI 傾向分析・コース推奨



質問や問合せの回答にAI 自動応答・QA対応
経済活動にAI 書類・技術・医療等

- 社会におけるAIの活用事例
- AIは万能か? (AIの得意・不得意)
- AI活用による新たな社会的課題
- AIとの適切な付き合い方

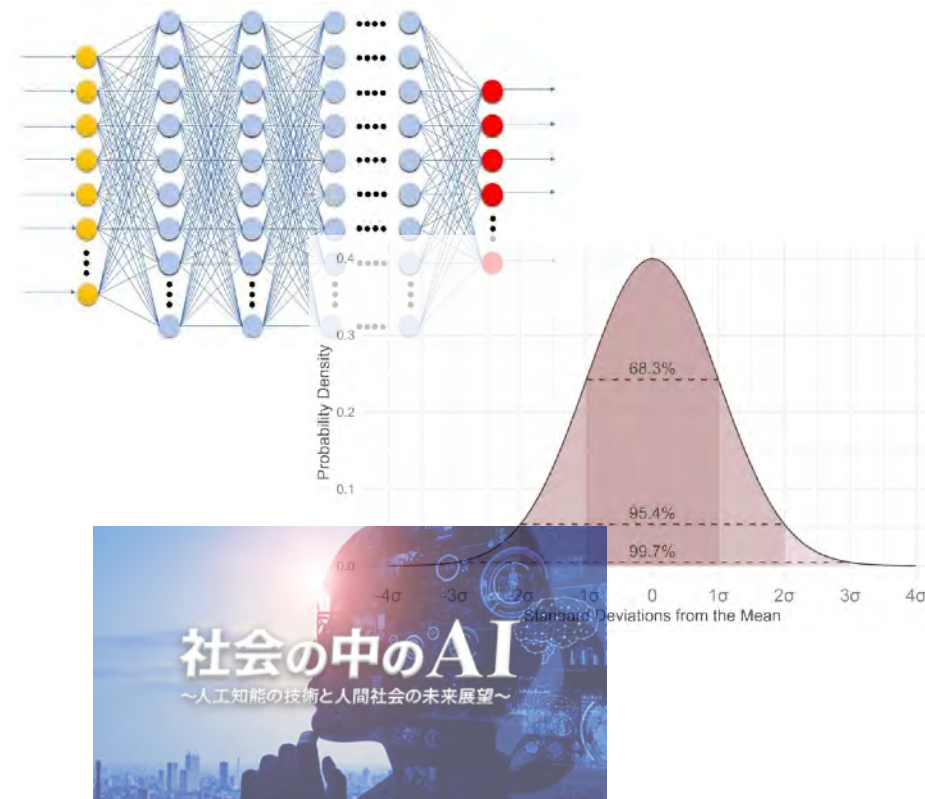
(Computational Data Science Program)

数理・データサイエンスの基礎的な手法を様々な実データに適用することで可能となる**問題解決スキル**を修得する

トピックス

- 機械学習や人工知能技術の中で用いられている基本的なアルゴリズムの理解と実践
- それらを支える数学的・統計学的な理論に対する理解
- 実社会への様々な応用事例に触れる体験
- プログラミングを伴う実践的なデータ処理演習

- ▶ **エキスパートレベル**へ進む基礎体力養成
- ▶ 理論・技術的側面だけでなく、**社会的・実務的視点**もカバー
- ▶ 一部科目を**学都仙台単位互換ネットワーク**を通して近隣大学・高専等に開放 (高大連携事業を通して高校生に公開されている科目もあり)



情報の基礎

機械学習の基本的なアルゴリズムおよび統計処理の理論的理解と実践

- 機械学習アルゴリズム概論
- 実践的機械学習I・実践的機械学習II
- データ科学・AI概論

AI等データ科学技術の社会現場

- 情報教育特別講義
 - AIをめぐる人間と社会の過去・現在・未来
 - AI・データ科学研究の現場

必修科目2つ (青字) + 4単位以上

統計の基礎

統計学の数学的基礎と数理モデリング

- 数理統計学入門／概論
- 数理統計学
- 情報教育特別講義
 - 統計数理モデリング

2単位以上

数学の基礎

データ科学・AIに関連する基礎的な数学に関する理論的理解

- 線形代数学入門／概論
- 線形代数学A／B
- 解析学入門／概論
- 解析学A／B

2単位以上

データサイエンス基礎

データエンジニアリング基礎・AI基礎

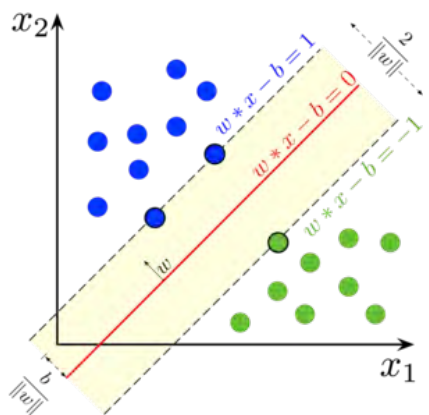
授業内容例: 実践的機械学習I (必修科目)

AIを支える基盤技術である機械学習について, そこに使われる基本的な考え方・アルゴリズムを実際に手を動かしながら (実践的に) そのエッセンスを学ぶ. (Python に関する高度な知識は前提としない.)

基本的な機械学習アルゴリズムの理論的理解

+

実データセットを用いた実践



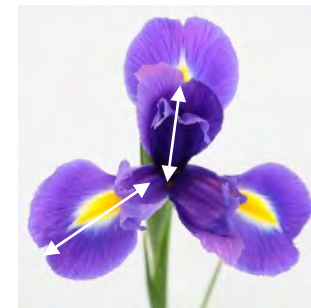
- ▶ 基本的な機械学習モデル
- ▶ データ分類・パターン認識の基本的な考え方

具体的な内容の事例

決定木, ニューラルネット, サポートベクトルマシン, 近傍法による分類など

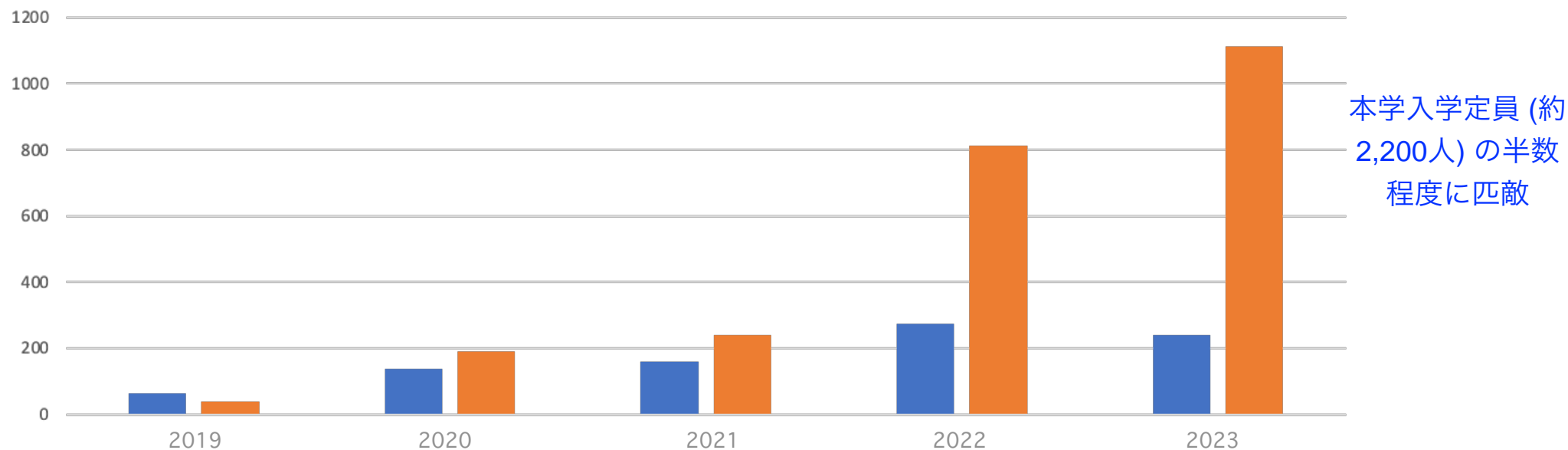
具体的な実習事例

- 花の画像を何種類かの特徴量をもとに種別別に分類するアルゴリズムを Python (Google Colaboratory 環境) を用いて実装
- 手書き文字データを利用して, データ分類器の動きを確認する, 次元削減の効果を確認する.



Scikit-learn, Numpy など機械学習ライブラリの使い方にも慣れる!

CDSプログラム必修科目履修状況 (2023年度)



本学入学定員 (約
2,200人) の半数
程度に匹敵

挑創カレッジCDS
プログラム提供開始

「情報基礎」にデータ
AI関連内容を追加
AIMD基礎を全学生必修に

現行「情報とデータの
基礎」開始

CDSプログラム 科目履修状況 (学部別 2023年度)

文系学生も積極的に履修

分類	授業科目名 <旧カリ名称>	CDS 修了 要件	学部 入学 定員	学部																			合計	その他		講義クラス数 (1学期/2学期)			
				文	教	法	経	理数	理物	理化	理地	理生	医医	医保	医保	医保	歯	薬	工機	工電	工化	工材		工建	農		院生	非正 規生	
A. 情報	実践的機械学習 I <実践 機械学習 1>	必修		7	7	1	28	7	34	18	6	3	5		11	1	1		12	48	12	33	1	4	239	12		1 (1:0)	
	機械学習アルゴリズム概論			54	9	29	101	35	59	43	39	21	7	2	8	2	2	7	199	194	87	112	94	8	1112	8	1	1 (0:1)	
	実践的機械学習 II <実践 機械学習 2>			1	1		5	2	12	7	1	2						11	17	2	1		3	67	1		1 (0:1)		
	情報とデータの基礎 <情報基礎A/情報基礎B>			241	79	171	284	46	130	77	50	43	123	73	38	39	58	87	236	255	118	121	111	167	2547	1	2	2 (2:1)	
	情報教育特別講義 (AIをめぐ る人間と社会の過去・現在・ 未来) <AIをめぐる人間と社 会の過去・現在・未来>	選択 (4単位)			8	5	13	12	2		1	1	2	1	3				12	31	7	19	5		122			1 (0:1)	
B. 統計	数理統計学入門 <統計学入門/数学概論D>	選択 (2単位)		30	60	3	86																		252			2 (1:1)	
	数理統計学概論 <数理統計学概要>				26		12				3	98		8	40	61	8							159	415			5 (4:1)	
	数理統計学						3	46	126	42	46	28	1					39	268	17	27	33	5	681	2	3	9 (8:1)		
	情報教育特別講義 (AI・デー タ科学研究の理論)			1			2	1			1	1	1	1	32			2	1	3	11	2		59			1 (1:0)		
C. 数学	解析学入門 <数学概論A>	選択 (2単位)		31	12	8	101																		154			2 (2:0)	
	線形代数学入門 <数学概論B>			28	3	14	2																		76			2 (0:2)	
	解析学概論 <数学概論C/解析学概要>						1					15	10	4	36		58	2							164	290			5 (0:5)
	線形代数学概論 <線形代数学概要>						59					4	53		38	37	1	80							177	449			5 (5:0)
	解析学A						26	46	131	85	49	41						1	253	265	118	133	125		1273			1 (6:1)	
	解析学B						19	50	130	77	41	21							262	265	129	135	115	1	1245			1 (6:1)	
	線形代数学A						62	47	128	85	50	46							244	259	119	127	116		1283	1		1 (6:1)	
	線形代数学B						42	49	128	79	47	20							199	258	115	117	115		1169			1 (5:1)	

「情報とデータの基礎」
(全学必修)

学部・学科の基礎科目という事情から履修率が高い

CDSプログラム修了認定状況

年度	学期	文	教	法	経	理	医	歯	薬	工	農	計	年間計
2020 (R2)	後期				2	2				1		5	5
2021 (R3)	前期				1					1		2	10
	後期		1		2					4	1	8	
2022 (R4)	前期				2				1	5	1	9	16
	後期				2	2				2	1	7	
2023 (R5)	前期		1			2			1	6		10	10
	後期											0	
累計		0	2	0	9	6	0	0	2	19	3	41	41

▶ 2020年度以降は修了者にはオープンバッジを発行



補足: 民間AI検定学修支援プログラムの状況

▶ 民間AI検定 (E検定・G検定) 学修支援プログラムは概ね毎年30名程度参加, 合格者は ~10名/年程度で推移



AIMD OpenCourseware

- ▶ 数学, 統計, プログラミングなどの自修教材 (動画も含む) をオンラインで提供
- ▶ コンテンツのさらなる充実を目指して目下活動中



教科書シリーズ「探検データサイエンス」の刊行

- ▶ センター教員も含む学内・学外の教員による教科書作成を継続中



AIMD課外活動・イベントの開催

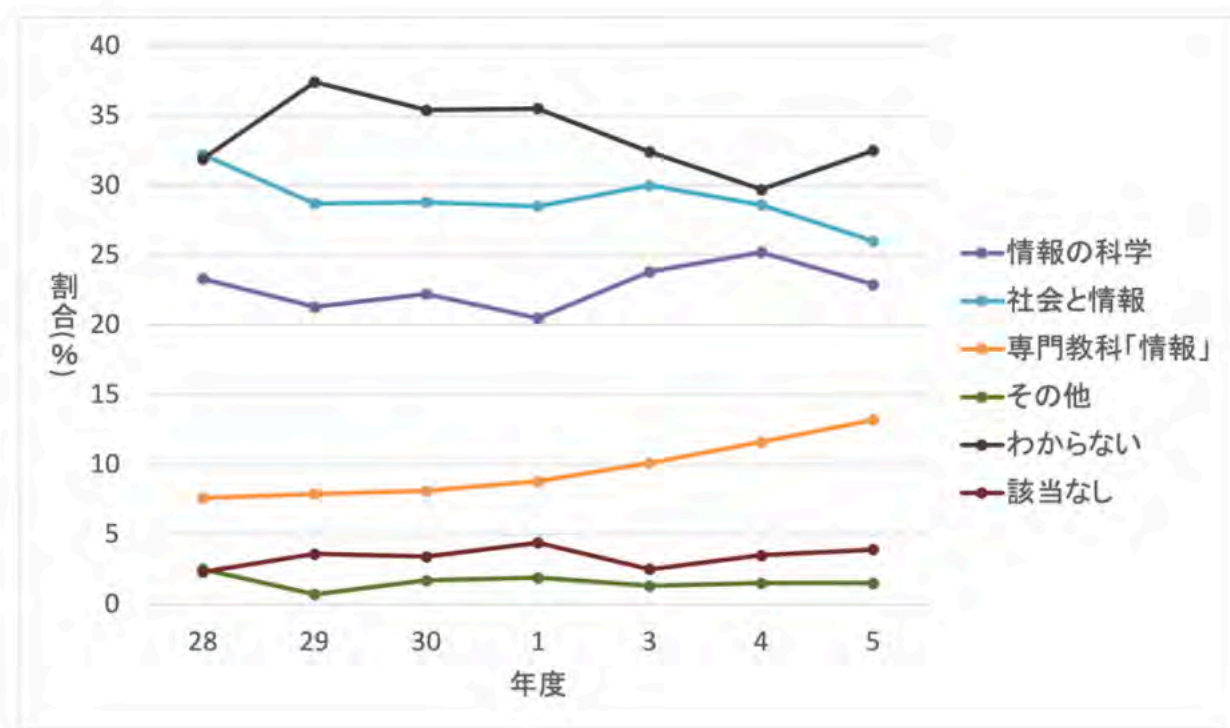
- ▶ 学部を問わず全学生を対象とした授業科目外での積極的な学びの場を提供

オンライン教材の開発・提供

- ▶ 企業や外部機関との共同開発による教材も含めて, AIMD関係の充実した教材群の開発・改良を継続中

2006年度 (平成18年度) から, 東北大学の新生を対象として, 「高等学校で履修した情報科目とそこで学習した内容」 および 「入学時の情報リテラシーレベル」 に関するアンケート調査を継続実施中.

近年の全体的な傾向



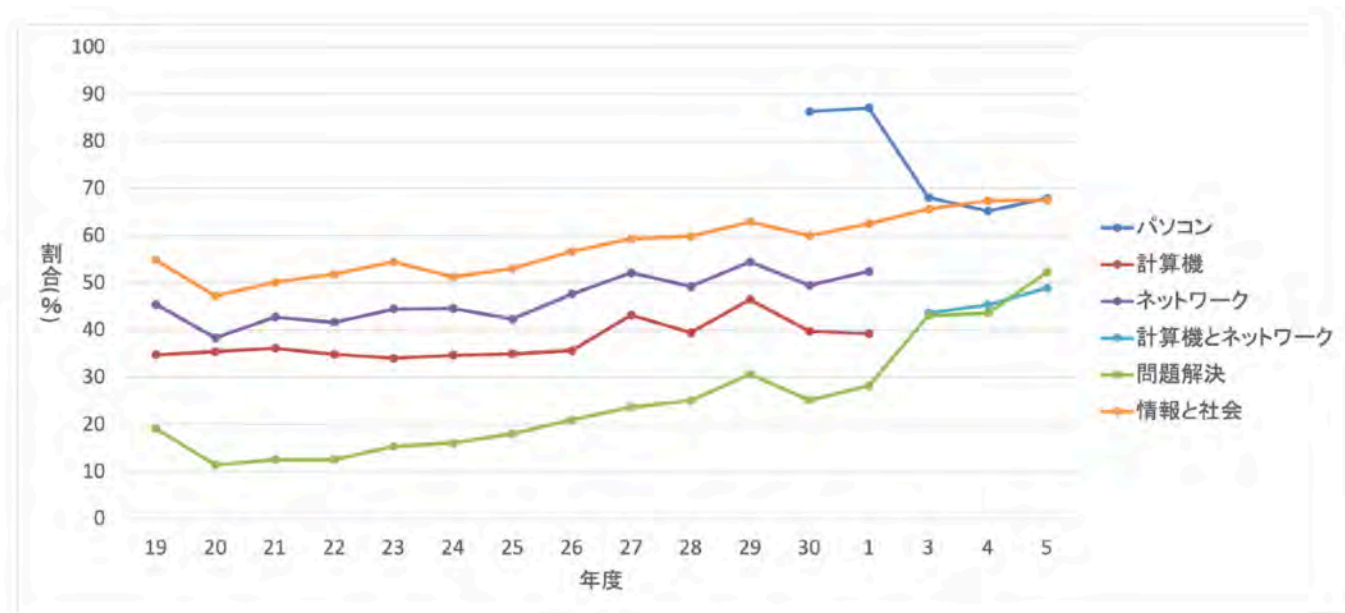
現役生の教科「情報」の履修状況の推移

① 教科「情報」の履修状況について

- 「情報の科学」と「社会と情報」がそれぞれ25%~30%ぐらいで推移している.
- 専門教科「情報」を履修した学生の割合が増加傾向にある.
- **自分が履修した科目がわからない学生が30%強存在する.**

2006年度 (平成18年度) から, 東北大学の新生を対象として, 「高等学校で履修した情報科目とそこで学習した内容」 および 「入学時の情報リテラシーレベル」 に関するアンケート調査を継続実施中.

近年の全体的な傾向



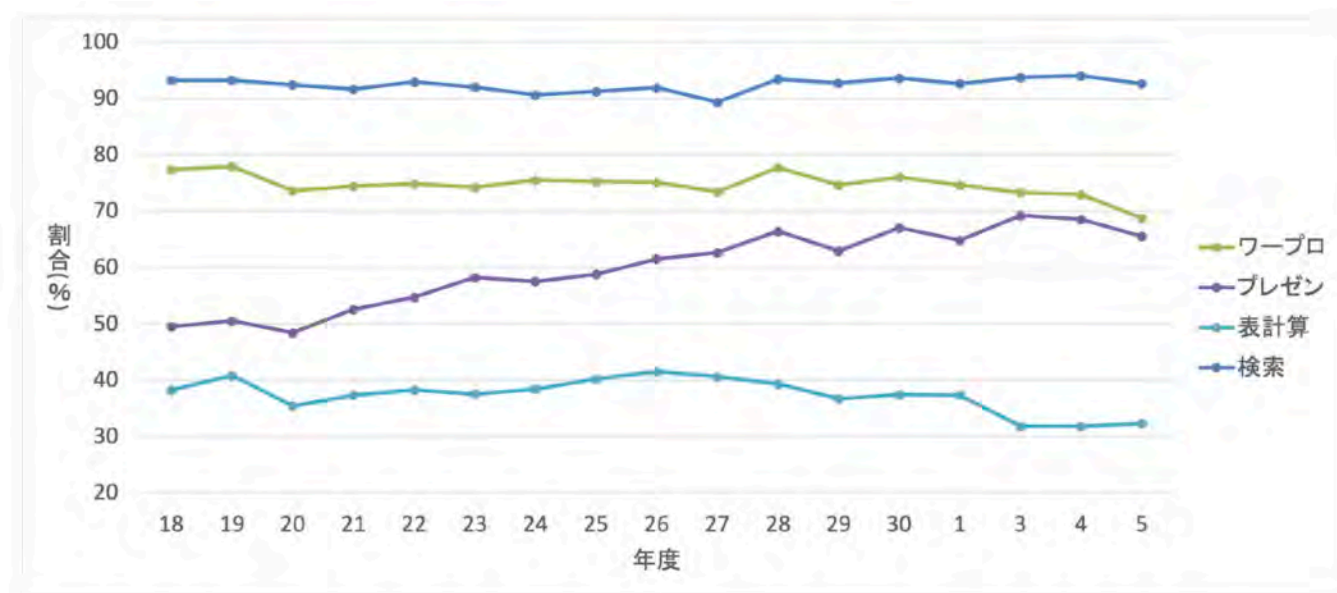
教科「情報」で学習した内容について
 (「基本的なことを学習した」 + 「詳しく学習した」の合計値)

② 「情報」で学習した内容について

- PCを用いた実習の割合が2021年度に大きく減少し, その後は横ばい傾向
 - ▶ 2020年度までは「PCの**使い方**実習」の経験について質問していたものが、2021年度以降は「PCを用いた**統計処理**実習」の経験についての質問に変えたことによる変化
- 2019年度以降, 「問題解決」の割合が明らかに増加傾向にある

2006年度 (平成18年度) から, 東北大学の新生を対象として, 「高等学校で履修した情報科目とそこで学習した内容」 および 「入学時の情報リテラシーレベル」 に関するアンケート調査を継続実施中.

近年の全体的な傾向



情報リテラシーレベルの自己評価について
(「まあまあできる」 + 「できる」の合計値)

③ 情報リテラシーレベルの自己評価について

- ここ20年近く, 全体的な傾向に大きな変化は見られない.
 - ▶ プレゼンテーションの増加割合が飽和しはじめている?
 - ▶ 表計算は例年低い水準で推移しているが, 表計算ツールのマクロを利用した実習を経験した学生などが増えてくれば傾向が変わる可能性があり, 経過観察が必要.



ご清聴ありがとうございました。