

高等学校における機械学習についての 指導の可能性と授業実践

埼玉県立川越南高等学校
春日井 優

自己紹介

- 春日井 優
- 高等学校 情報科教諭
- 数学で高校教諭に採用
- 2000年の現職教員講習で情報免許を取得
- 情報科専任になって5年
- 勤務校は大学進学する生徒が大多数

発表内容

- 背景
 - 社会の変化
 - 学習指導要領の変化
- 高等学校での機械学習の指導の可能性
- 高等学校で機械学習を指導した一事例の紹介
- 今後の課題

高等学校で機械学習を指導する背景

- 社会の状況
- 学習指導要領の変化

社会の状況①

- Watson

VS

米人気クイズ番組

歴代チャンピオン2人

Watsonが圧勝

(2011年)

社会の状況②

- 将棋ソフトがプロ棋士に勝利 (局面数 10の220乗)

あから2010

VS 清水女流王将
(2010年)

電王戦5局

コンピュータ3勝1敗1分
(2013年)

ponanza

VS 佐藤天彦名人
(2017年)

社会の状況③

- AlphaGoがプロ棋士に勝利 (局面数 10の360乗)

VS イ・セドル九段
AlphaGo 4勝1敗
(2016年)

VS 柯潔九段
AlphaGo 3勝
(2017年)

社会の状況④

- 東口ボくん

偏差値57.1

A判定(合格可能性80%)

私立大 512大学

国公立大 23大学

社会の状況⑤

- ・ マイケル・オズボーン「雇用の未来」

米国の労働人口の47%
10～20年に
機械に代替されるという
推計

社会の状況⑥

- スマートスピーカーなど
人工知能を謳う製品の普及

AIスピーカー

AI家電

AI自動車

AI先生

社会の状況⑦

- 行政もAI頼り

AIによる

- ・工場内故障診断
- ・医療機器等の開発
- ・梨摘果判断
- ・プッシュ型サービス
- ・救急相談
- ・結婚支援

まとめると

- 人間のトップより人工知能ってすごいね！
- もちろん大抵の人間よりも能力高いよ！
- 人間の仕事がなくなるかもね...
- AIが身近に使えるようになってきた

← 今、このあたり？

- AI様をお願いしてみよう

..... コンピュータが自動で処理をしたり、
知識を得て回答をする仕組みを
専門家以外も知っている必要があるのでは？

高校で機械学習を指導する背景

- 社会の状況
- 学習指導要領の変化

小学校・中学校・高校では どうなっている？

- 現行の学習指導要領
 - 小学校・中学校 2008年改訂
 - 高校 2009年改訂

まだ、この頃は人工知能は遠い未来の話.....

何も人工知能についての記述はありません！

次の時代の学習は？

2016年 中教審答申で方向性が示される

2017年 小学校・中学校学習指導要領改訂
解説で学習の詳細が示される

2018年 2月14日

高等学校学習指導要領案

現在、パブリックコメント募集中

2018年 高等学校学習指導要領改訂

解説も発行されるはず

中教審答申

- 2030年を見据えて
 - 知識基盤社会 → 知識・情報・技術の変化
 - 人工知能の進化 → 雇用環境の変化
学校で学ぶことが通用しない？
 - 情報技術の進化 → グローバル化
 - 予測不能な社会
 - 人工知能にはない 豊かな感性, 目的をもつ,
場面に合わせた納得解

小学校・中学校の 学習指導要領解説では

- 人間の強み
 - 思考の目的を与えたり, 目的のよさ・正しさ・美しさを判断できる

人工知能と比較して, 人間の役割やよさが書かれている

小学校の学習内容

- 人工知能の記述はありませんでした

中学校の学習内容①

技術科(情報の技術の単元)

「生活や社会における人工知能の活用について、人間の労働環境や安全性、経済性の視点から、その利用方法を検討する」

中学校の学習内容②

社会科(公民分野)

「情報化については、

人工知能の急速な進化などによる

社会の構造的な変化などと関連付けたり、

災害時における防災情報の発信・活用などの

具体的な事例を取り上げたりすることとした」

中学校の学習内容③

理科(第1分野「科学技術と人間」)

「科学技術の発展について

科学技術の発展の過程を知るとともに、
科学技術が人間の生活を豊かで
便利にしていることを認識すること。

例えば, ...人工知能...など

最新の科学技術を調べさせることが考えられる」

2月14日に示された 高等学校学習指導要領案①

- 世界史探究(「科学技術の高度化と知識基盤社会」)

「(前略), **人工知能と労働の在り方の変容**,
情報通信技術の発達と知識の普及などを基に,
知識基盤社会の展開と課題を理解すること。」

2月14日に示された 高等学校学習指導要領案②

- 世界史探究(「現代の諸課題と倫理」)
「生命, 自然, 科学技術などと人間との関わりについて
倫理的課題を見いだし...

「科学技術」については,

...人工知能(AI)をはじめとした先端科学技術の利用と
人間生活や社会の在り方についても思索できるよう
指導すること」

人工知能が含まれる箇所は, 以上でした

これまでの議論から推測すると

- 人工知能が進化し普及していく社会を想定している
- 人工知能の活用事例を知る,
技術について調べる,
人間や社会の在り方について考える

情報活用の実践, 情報社会に参画する態度

情報の科学的な理解が不十分では？

もう少しよく読むと

- 数学科と情報科で統計の充実
 - 統計を基に考えられた仕組みもある
- 情報Ⅱでは「データサイエンス」を扱う単元が新設される

情報 I の内容案

「情報 I」(情報と情報技術を問題の発見と解決に活用するための科学的な考え方等を育成する共通必修科目)

問題の発見・解決に向けて、事象を情報とその結び付きの視点から捉え、情報技術を適切かつ効果的に活用する力を育む科目

(項目の構成案)

(1) 情報社会の問題解決	中学校までに経験した問題解決の手法や情報モラルなどを振り返り、これを情報社会の問題の発見と解決に適用して、情報社会への参画について考える。
(2) コミュニケーションと情報デザイン	情報デザインに配慮した的確なコミュニケーションの力を育む。
(3) コンピュータとプログラミング	プログラミングによりコンピュータを活用する力、事象をモデル化して問題を発見したりシミュレーションを通してモデルを評価したりする力を育む。
(4) 情報通信ネットワークとデータの利用	情報通信ネットワークを用いてデータを活用する力を育む。

情報Ⅱの内容

「情報Ⅱ」(発展的な内容の選択科目)

「情報Ⅰ」において培った基礎の上に、問題の発見・解決に向けて、情報システムや多様なデータを適切かつ効果的に活用し、あるいは情報コンテンツを創造する力を育む科目

(項目の構成案)

(1) 情報社会の進展と情報技術	情報社会の進展と情報技術との関係について歴史的に捉え、AI等の技術も含め将来を展望する。
(2) コミュニケーションと情報コンテンツ	画像や音、動画を含む情報コンテンツを用いた豊かなコミュニケーションの力を育む。
(3) 情報とデータサイエンス	データサイエンスの手法を活用して情報を精査する力を育む。
(4) 情報システムとプログラミング	情報システムを活用するためのプログラミングの力を育む。
○ 課題研究	情報Ⅰ及び情報Ⅱの(1)～(4)における学習を総合し深化させ、問題の発見・解決に取り組み、新たな価値を創造する。

情報社会の進展と情報技術

- 知識

(ウ) 情報技術の発展による人の知的活動への影響について理解すること。

- 思考・判断・表現

(ウ) 人の知的活動が変化する社会における 情報システムの創造やデータ活用の意義について考察すること。

情報とデータサイエンス①

●知識

- (ア) 多様かつ大量のデータの存在やデータ活用の有用性、データサイエンスが社会に果たす役割について理解し、目的に応じた適切なデータの収集や整理、整形について理解し技能を身に付けること。
- (イ) データに基づく現象のモデル化やデータの処理を行い解釈・表現する方法について理解し技能を身に付けること。
- (ウ) データ処理の結果を基にモデルを評価することの意義とその方法について理解し技能を身に付けること。

情報とデータサイエンス②

●思考・判断・表現

(ア) 目的に応じて、適切なデータを収集し、整理し、整形すること。

(イ) 将来の現象を予測したり、複数の現象間の関連を明らかにしたりするために、適切なモデル化や処理、解釈・表現を行うこと。

(ウ) モデルやデータ処理の結果を評価し、モデル化や処理、解釈・表現の方法を改善すること。

データサイエンスと人工知能と機械学習

ここからは

- 人工知能

コンピュータが行う認知や推論などの
知的な情報処理

(単に自動制御のようなものでも

人工知能と呼んでいる人がいるので...)

範囲を機械学習に限定

回帰による予測, 分類, クラスタリング

高校での機械学習の指導の可能性

- 高校での学習内容
- 機械学習との対応

高等学校において指導可能とは

- 高等学校で学習した内容の発展にあたるもの
例 分散・共分散の求め方 → 値を用いる
- 高等学校で学習する内容と同程度の概念のもの
例 脳細胞が信号を伝達する仕組み
≡
人工ニューロンが信号を伝達する仕組み

とした

関係する内容(情報科・現行)

- 「社会と情報」
 - 仕組みに関しては特にない
- 「情報の科学」
 - モデル化・処理手順の自動化
 - データを扱うためのモデル
 - 処理するためのアルゴリズム
 - 処理の自動化

関係する内容(情報科以外・現行)

- データの分析(数学Ⅰ)
- 図形と方程式(数学Ⅱ)
- 場合の数と確率(数学A)
- 確率分布と統計的な推測(数学B)

- 生物の環境応答・刺激の受容と反応(生物)

高校での機械学習の指導の可能性

- 高校での学習内容
- 機械学習との対応

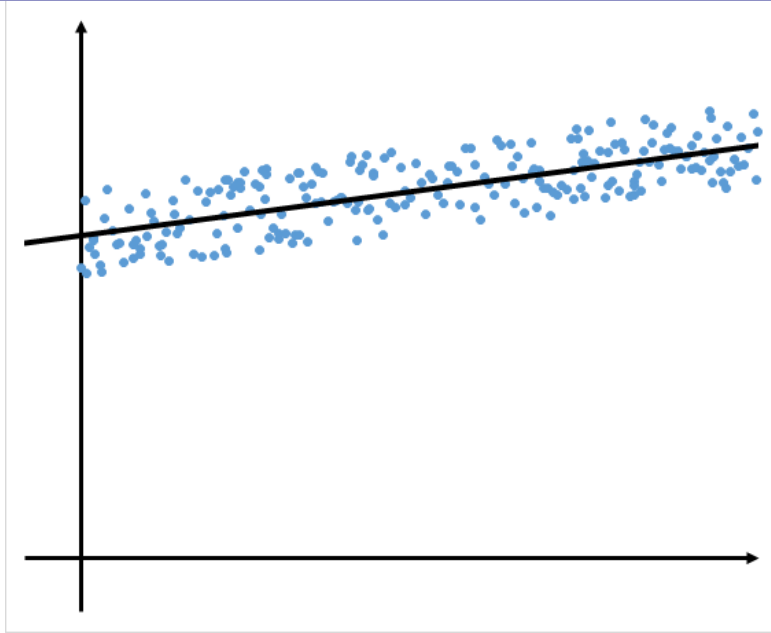
検討の対象とする機械学習

- 広く高校生が学習する内容
最先端ではなく, 広く一般化した内容

今回, 検討の対象とした機械学習

- | | |
|---------------|--------|
| ① 回帰分析 | ② k近傍法 |
| ③ サポートベクターマシン | ④ 決定木 |
| ⑤ 単純ベイズ分類器 | ⑥ k平均法 |
| ⑦ ニューラルネットワーク | |

① 回帰分析



概念の理解

1次関数に近似
(中学校数学)

アルゴリズムの理解

最小二乗法

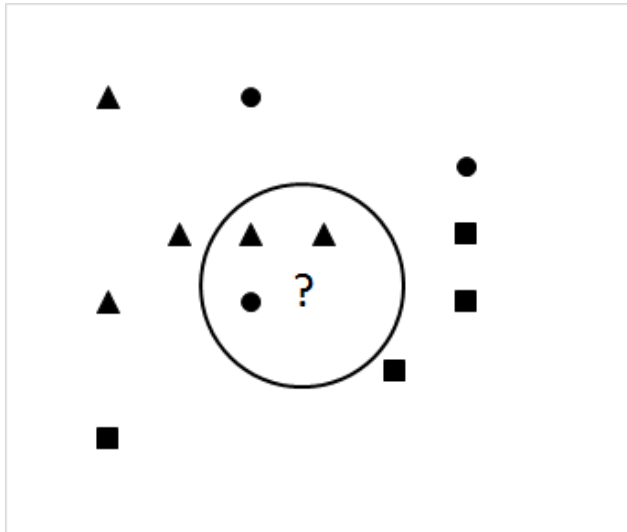
求めるには分散・共分散が
必要(数学 I)

$$a = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b = \bar{y} - a\bar{x}$$

② k近傍法

概念の理解



距離の概念

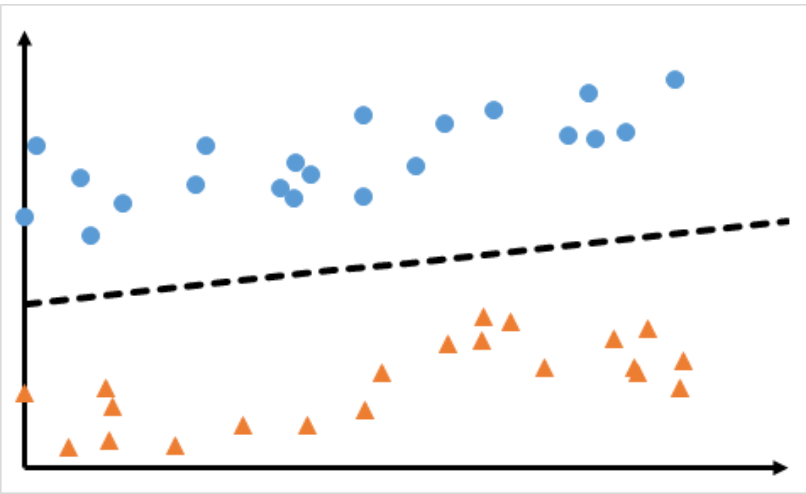
アルゴリズムの理解

ユークリッド距離を求める

三平方の定理(中学校数学)

2点間の距離(数学Ⅱ)

③ サポートベクターマシン



概念の理解

不等式が表す領域(数学Ⅱ)

アルゴリズムの理解

ラグランジュの未定乗数法
(高校の範囲外)

④ 決定木

概念の理解

条件分岐

アルゴリズムの理解

確率(数学A)

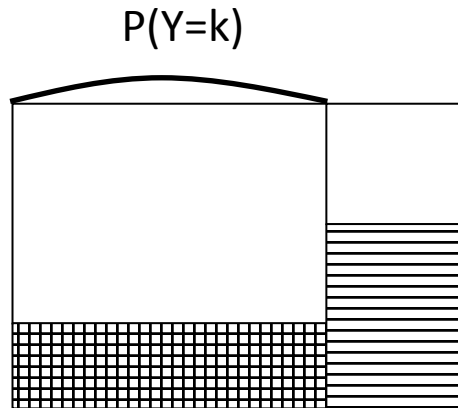
$p(j|t)$ は相対頻度


$$I_{\square}(t) = 1 - \sum_{j=1}^c p(j|t)^2$$



⑤ 単純ベイズ分類器

概念の理解



 $P(X|Y=k)$

 +  = $P(X)$

条件付き確率 (数学A)

アルゴリズムの理解

実装するには

アンダーフロー対策が必要

その際に対数関数が必要

(数学II)

⑥ k平均法

概念の理解

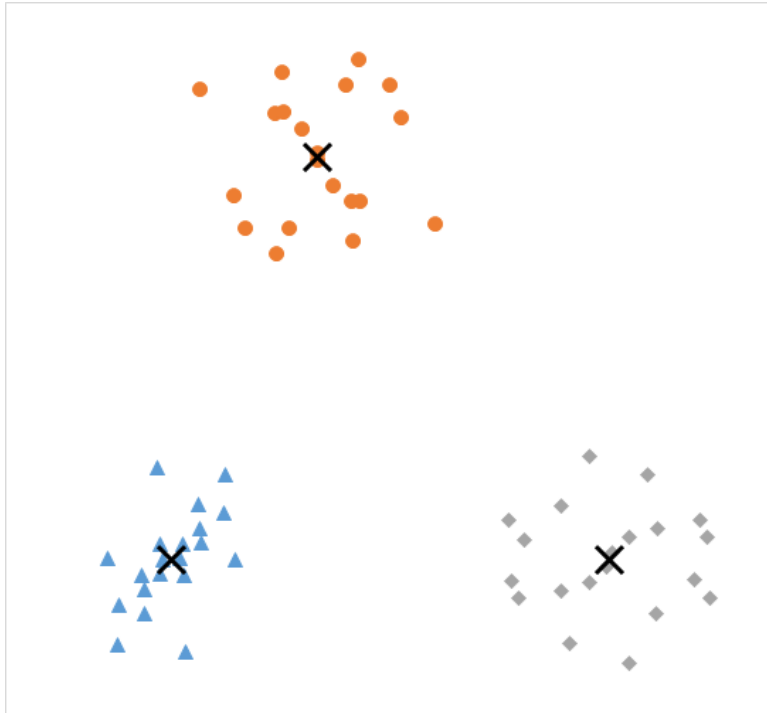
距離の概念・平均値

アルゴリズムの理解

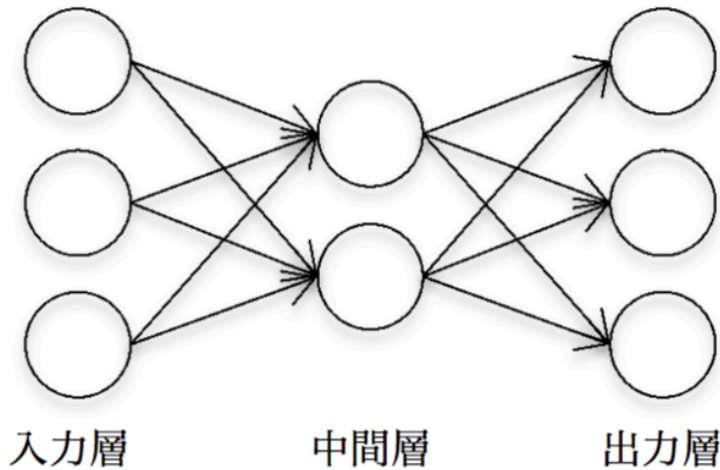
平面上の2点間の距離

(数学Ⅱ)

(三平方の定理は中学校数学)



⑦ ニューラルネットワーク



概念の理解

神経細胞(脳細胞)での
信号の伝達(生物)

アルゴリズムの理解

活性化関数が
高等学校の範囲外

まとめると

アルゴリズム	概念の理解	アルゴリズムの理解
①回帰分析	1次関数	分散・共分散
②k近傍法	距離	2点間の距離
③サポートベクターマシン	不等式が表す領域	×
④決定木	条件分岐	確率
⑤単純ベイズ分類器	条件付き確率	対数関数
⑥k平均法	距離	2点間の距離
⑦ニューラルネットワーク	信号の伝達	×

概念・アルゴリズムについて
 高等学校の知識で対応可能なものが多い

高等学校で機械学習を指導した 一事例の紹介

高等学校における授業実践

- 勤務校での実践
- 3学年「情報の科学」
モデル化とシミュレーション として実践

3学年で開講していることから

数学 I・II・A は全員履修済み

数学B は選択者のも履修済み

授業内容

単純ベイズ分類器の仕組みの指導

理解を深めるため

TF-IDF法での特徴語抽出

ワードクラウドでの可視化

TF-IDF法・単純ベイズ分類器を活用できる

問題の発見→解決(案)の提案

具体的な例が少ないので、
教材の開発をしていくことが必要になる

単純ベイズ分類器

使われる語(形態素)の
数を数えて
紙と鉛筆で求める



アンダーフロー対策の必要性



プログラムの該当箇所の確認

```
for category in word_list:
    score = math.log( count_documents[category] / count_all )
```

カテゴリーの数

```
# (2)分類内単語出現率のlogをscoreに加えることで(3)の文章内分類出現率を求める
score += math.log( ( word_list[category][word] + 1 ) / ( sum + len(words) ) )
```

☆ 単純ベイズ分類器

ベイズの定理の考えを用いて、機械的に分類してみよう。

学習データ

いちご : [果物 | ケーキ | ビタミン]
りんご : [果物 | ジュース | ケーキ | ビタミン | 赤い]
キウイ : [ビタミン | 毛 | 緑 | 黄色]
いちご : [ケーキ | 赤い]

分類データ

: [果物 | ケーキ]

1) 分類出現率の計算 ($P(\text{category}) = \text{文章数} / \text{全文章数}$)

	いちご	りんご	キウイ
分類出現率	2/4	1/4	1/4

2) 分類内の単語出現率 ($P(\text{word} | \text{category}) = \text{単語数} / \text{分類の総単語数}$)

	いちご	りんご	キウイ
単語数	5	5	4
果物	1/5	1/5	0/4
ケーキ	2/5	1/5	0/4
分類内の 単語出現率	$(1/5) \times (2/5)$ = 2/25	$(1/5) \times (1/5)$ = 1/25	$(0/4) \times (0/4)$ = 0/16

3) 文章内の分類出現率 ($P(\text{category} | \text{word}) = P(\text{category}) \times P(\text{word} | \text{category})$)

	いちご	りんご	キウイ
文章内の 分類出現率	$(2/4) \times (2/25)$	$(1/4) \times (1/25)$	$(1/4) \times (0/16)$
計算結果	4/100	1/100	0

以上の計算結果より、最も確率が高いのは いちご に分類されることになる。

TF-IDF法

- 特定のカテゴリーに頻出する語(形態素)が値が高くなる

TF-IDF法

単純ベイズ分類器

TFが高い

→

確率が高くなる

IDFが高い

→

分類されるカテゴリーだけ
確率が高くなる

傾向がある

分類されるカテゴリーの予測に使える

ワードクラウド

- 頻出語の文字が大きい
大きい
↓
特定の 카테고리で
大きい語(形態素)を
使うと
その 카테고리に分類されそう



分類されるカテゴリの予測に使える

問題発見・解決のために 生徒に使わせた

- 客の希望をもとに，希望に合った観光地を示す
- 好みの性格や特徴をもとに，おすすめの異性をタレントに例えて示す
- 知りたい情報をもとに，
朝のテレビ番組の中からおすすめの番組を示す
- アレルギーの症状をもとに，原因となっている花粉を示す
- 好みの犬の性格をもとに，おすすめの犬種を示す
- 食べたい料理その調理法をもとに，その料理にあう
サツマイモの品種を示す

など

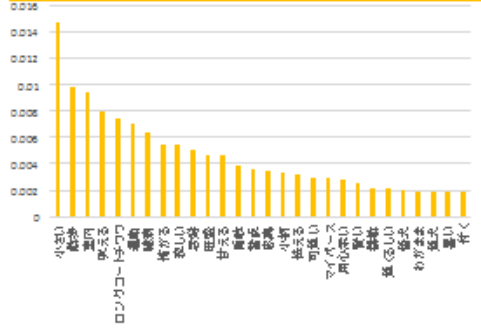
生徒の成果物

チワワの特徴語

ワードクラウド



Tf-idf



ナイーブベイズによる お客さんの要望に対する回答

- 「賢く活発で可愛い犬が欲しいです」
ポメラニアンがおすすめ
- 「小さくて忠実で甘えん坊の犬が欲しいです」
チワワがおすすめ
- 「人気で飼いやすくて従順な犬が欲しいです」
柴犬がおすすめ

授業後の生徒のコメント

- 分類に**必要な単語が多く含まれる**文章を集める必要がある
- **必要でない単語**を消す必要がある
- **キーワード**になっている言葉が多く含まれるものに分類される
- **似た単語**が使われているものの分類は難しい
- **データの量**(事前確率に使われている行数)の差で、結果が大きく異なる
- アンケートなどをもとに分類するとき、**ふざけた回答があると影響を受ける**

授業の成果

- 高等学校においても機械学習の授業ができることがわかった

生徒

- 仕組みを知ることによってコンピュータが出す回答をある程度予想できるようになった
- 仕組みを知ることによって誰かが書いたプログラムで動いていることがわかった
- 機械学習での、データの重要性がわかった

今後の課題

今後の授業内容として検討すべきこと

- データの重要性
 - 悪意があるデータの除去
(マイクロソフト Tay)
- 著作権
 - 著作権法第47条7「情報解析のための複製」
 - コンピュータが作り出したものの著作権は？

他教科(数学)との関係①

- 距離の計算の拡張

幾何的には3次元が限界だけど...

4次元以上に拡張すると
いろいろできるけど...

$$\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

他教科(数学)との関係②

• ベクトルの扱い(次期学習指導要領案)

知識・技能

- (ア) 平面上のベクトルの意味, 相等, 和, 差, 実数倍, 位置ベクトル, ベクトルの成分表示について理解すること。
- (イ) ベクトルの内積及びその基本的な性質について理解すること。
- (ウ) 座標及びベクトルの考えが平面から空間に拡張できることを理解すること。

思考・判断・表現

- (ウ) 数量や図形及びそれらの関係に着目し, 日常の事象や社会の事象などを数学的に捉え, ベクトルやその内積の考えを問題解決に活用すること。

