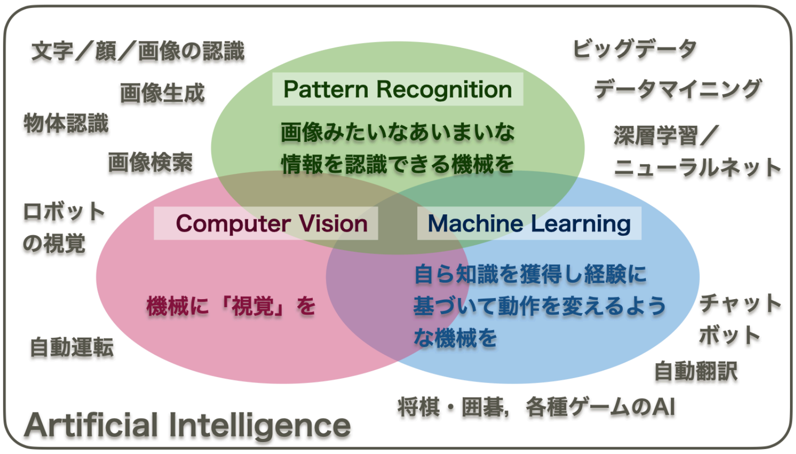
目次

★ 10　パターン認識と機械学習 (1) — パターン認識，機械学習とは

今回から，**パターン認識** (☆ 1) と**機械学習** (☆ 2) を取り上げる．以下の図は，これら 2 つの分野と関連の深い**コンピュータビジョン** (☆ 3) を加えて，これらの分野とその周辺のキーワード（のうちごく一部）を示したものである．

☆ 1) パターン認識: pattern recognition
☆ 2) 機械学習: machine learning
☆ 3) コンピュータビジョン: computer vision



★ 10.1　パターン認識とは

パターン認識とは，コンピュータを用いて，画像や音声のようなパターン情報の中の規則性を自動的に見つけ出し，それに基づいてパターンをいくつかのカテゴリに分類するような情報処理のことをいう（第 1 回講義資料も参照のこと）．音声認識，文字認識，画像認識などが典型例である．例えば...

例 PR1: 母音の音声波形から「a,i,u,e,o」を識別する (波形を ‘a’, ‘i’, ‘u’, ‘e’, ‘o’ の 5 つのカテゴリに分類する)

例 PR2: 人の顔画像をいくつかのカテゴリに分類する (カテゴリの設定の仕方次第で様々な応用が考えられる)

例 PR3: 様々な動物の写った画像が与えられたときに，それぞれの画像に写っている動物の種類を識別したい (ネコ，ペンギン，カピバラ,⋯)

パターンをカテゴリ分類する処理とは少し異なるが，類似の情報処理として，次のような問題もパターン認識で扱う対象となる．

例 PR4: 顔画像から年齢を推定する

例 PR5: 大量の画像の中から，ある画像と似た画像を見つけ出す

近年のパターン認識では，研究レベルでも実用レベルでも，後述する機械学習を利用した方法を採ることがほとんどである．

★ 10.2　機械学習とは

機械学習とは，人間のように自ら知識を獲得し経験に基づいて動作を変えるような機械／コンピュータ／アルゴリズムの実現を目指す研究分野またはその技術を指す．人間が行なっているような知的な情報処理の仕組みを人工的に実現することを目指す**人工知能** (☆ 4) の一分野であるが，現在ではより幅広く，データの中から何らかの知識や法則性を自動的に見つけ出したり，データに基づいて何らかの判断を自動的に下したりするような情報処理全般を対象とするものとなっている．大量のデータが与えられたときに，そこに潜む規則性を自動的に見つけ出し，そこから何らかの情報を取り出すような情報処理は，機械学習の応用の例である．例えば... (以下の括弧の中は，取り出す情報がどのようなものかを表す)

☆ 4) 人工知能: artificial intelligence, AI

例 ML1 時々刻々変化する気温，湿度，電力使用量の数値から，少し未来の電力使用量の推定値を求めたい (未来の電力使用量 (実数値))

例 ML2 通販サイトの購買履歴データをもとに，顧客を買い物傾向の似たグループに分類したい (例えば「グループ 1」から「グループ K」までの K 通りのいずれか)

先のパターン認識の例も，機械学習の問題の一つと考えることができる．

例 PR1: 母音の識別 (‘a’, ‘i’, ‘u’, ‘e’, ‘o’ の 5 つのカテゴリのいずれか)
例 PR2: 顔画像の認識 (顔/非顔，男性/女性, A さん/B さん/C さん..., etc.)
例 PR3: 画像中の動物の認識 (ネコ/ペンギン/カピバラ/⋯)
例 PR4: 顔画像から年齢を推定する (年齢の数値)

これらの処理は共通して，何らかのデータ $\boldsymbol{x}$ が入力されると適当な値 $\boldsymbol{y}$ を出力する関数として次式のように表すことができる．

$$\boldsymbol{y} \longleftarrow \boxed{f} \longleftarrow \boldsymbol{x} \qquad\qquad \boldsymbol{y} = f(\boldsymbol{x};\boldsymbol{w}) \tag{1}$$

入力 $\boldsymbol{x}$ は，問題に応じて様々な形式で与えられる．出力 $\boldsymbol{y}$ (☆ 5) は，例 ML1 や PR4 のように多様な数値をとる場合もあれば，例 ML2 などのように何通りかの値しかとらない場合もある．ここで $\boldsymbol{w}$ は，関数 f に含まれるパラメータ (変数) (☆ 6) である (☆ 7)．パラメータを変えると，同じ入力に対しても異なる出力を与えることになる．このように式で表現すると，機械学習とは,「入力に対して「望ましい出力」が得られるように，パラメータ $\boldsymbol{w}$ をデータから自動的に決定する方法」に関する分野/技術と言うことができる．

機械学習において，望ましい出力が得られるようにパラメータを調節する過程を**学習** (☆ 8) といい，その際に用いるデータを**学習データ** (☆ 9) という．学習データで望ましい出力が得られることは当然大事であるが，機械学習の目標は，学習後に，学習時に遭遇したことのない未知のデータが入力されても，望ましい出力が得られる (これを「**汎化能力**がある」という) ようにすることである．

☆ 5) ここの例は全て出力が一つの値をとるものであるが，一般には出力が複数の値の組であってもよいので，ここでは $\boldsymbol{y}$ をベクトルして表している．

☆6) パラメータ: parameter, 助変数，媒介変数．

☆ 7) 例えば，$y = f(x) = ax + b$ では a と b がパラメータ．

☆8) 学習: learning, 訓練 (training) とも．

☆ 9) 訓練データともいう．

★ 10.3　機械学習のアルゴリズム・手法の分類

機械学習の方法は，学習データの与え方によって次の 2 つに大別される．

教師あり学習 (☆ 10) 個々の学習データが，入力とそれに対する出力の正解のペアとして与えられる．前述の例 PR4 では，顔画像 (入力) とその人の年齢 (出力の正解) のペアを複数集めたものを学習データとすればよいだろう．特殊な場合として，**強化学習** (☆ 11) を含む．

教師なし学習 (☆ 12) 学習データは入力のみで構成され，出力の正解は与えられない．どのような出力が望ましいかが学習アルゴリズムに反映されており，データの分布などに応じて出力が自己組織 (☆ 13) される．**自己組織学習**と呼ぶこともある．

例 ML1 や PR4 は，教師あり学習の典型例である．一方，ML2, PR1,2,3 の場合，学習データとしてどのような分類が正解かという情報を与える (「このお客さんはグループ 1」,「この顔画像は女性」，etc.) なら教師あり学習の問題であるし，入力データだけを与えて自動的に分類させるならば教師なし学習の問題となる．

☆ 10) 教師あり学習: supervised learning.

☆ 11) 強化学習: reinforcement learning. 出力の正解が直接与えられるのではなく，ある入力に対する出力が「良かった」「悪かった」というような評価の値のみが与えられる形態の学習．ロボットの行動の学習などによく用いられる．

☆ 12) 教師なし学習: unsupervised learning.

☆ 13) 自己組織: self-organization.

★ 10.3.1　教師あり学習

教師あり学習の問題は，式 (1) の出力 $\boldsymbol{y}$ がどのような値であるかによって，**回帰** (☆ 14) と**識別** (または**分類**) (☆ 15) の二つに大別できる．

回帰 出力として量的な値を扱う場合．例 ML1 と PR4 が該当する．

識別 入力をいくつかのグループ (**クラス**または**カテゴリ**という) に分ける問題．上記の 2 つの例以外の上述の例は，教師あり学習の問題として扱うならば全てこちらに入る．例えば例 PR1 なら ‘a’, ‘i’, ‘u’, ‘e’, ‘o’ の 5 クラスを識別する問題といえる．

☆ 14) 回帰: regression

☆ 15) 識別/分類: classification

★ 10.3.2　教師なし学習

教師なし学習の適用対象となる問題の代表例を以下にあげる．それぞれについて様々な学習アルゴリズムが存在する．

クラスタリング データをいくつかのかたまり (**クラスタ** (☆ 16)) に分類する．

特徴抽出/次元圧縮 高次元のデータから本質的な情報を抽出する，情報を保持したまま低次元に変換する．この授業でパターンの成分分析の手法として学んできたもの (ベクトルの直交展開，フーリエ変換等) と関連が深い．

頻出パターン抽出 データ中に一定の頻度以上で出現するパターンを見つけ出す．

これらは，大量のデータの中から有益な情報を見つけ出す**データマイニング** (☆ 17) の対象とする問題でもある．教師なし学習とデータマイニングという 2 つの分野は大きく重なっている．

☆ 16) クラスタ: cluster, クラスタリング: clustering.

☆ 17) データマイニング: data mining.　英語の動詞 mine には，(鉱石などを) 採掘する/掘り出すという意味がある．

★11　パターン認識と機械学習 (2) — 回帰のための教師あり学習

教師あり学習の目的は，学習データを用いて入力 $\boldsymbol{x}$ に対するシステムの出力 $\boldsymbol{y} = f(\boldsymbol{x}; \boldsymbol{w})$ が正解に近づくようにパラメータ $\boldsymbol{w}$ を調節し，未知の入力データに対しても望ましい出力が得られるようにすることである．ここでは，出力 $\boldsymbol{y}$ として量的な値を扱う問題，すなわち**回帰問題**を対象とし，その最も簡単な解法として，**最小二乗法**を取り上げる．

★11.1　最小二乗法による直線の当てはめ

回帰問題の最も簡単な形は，入力も出力も 1 次元で，入出力の関係として直線

$$y = f(x; a, b) = ax + b \tag{2}$$

を考える場合である．パラメータは a, b である．学習データとなる入力と出力の正解のペアが $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ の N 個与えられるとすると，この問題は，これら学習データに当てはまる直線を見つける問題といえる (図 1 参照)．つまり，$n = 1, 2, \ldots, N$ のそれぞれについて

$$y_n \approx ax_n + b \tag{3}$$

となるようにパラメータ a,b を定める問題である (☆ 18)．このように考えると，この問題を解くには，学習データに対する直線当てはめの「良さ」を定義して，最も良いパラメータ (a, b) を選べばよいことがわかる．このような回帰問題を含めたデータ解析の最も基本的な手法の一つである**最小二乗法** (☆ 19) では，この「良さ」の規準として，入力 x_n に対する出力 $f(x_n)$ とその正解 y_n との**二乗誤差** (☆ 20) を考える (☆ 21)．いま考えている直線当てはめ (直線回帰) の問題では，二乗誤差は

$$(y_n - f(x_n))^2 = (y_n - (ax_n + b))^2 \tag{4}$$

と表される．このとき，学習データに対する二乗誤差の和として誤差関数 $E(a, b)$ を次のように定義する．(☆ 22)．

$$E(a, b) = \frac{1}{2} \sum_{n=1}^{N} (y_n - (ax_n + b))^2 \tag{5}$$

$E(a, b)$ が最小となるような a, b を求めたいので，$\frac{\partial E}{\partial a} = 0, \frac{\partial E}{\partial b} = 0$ とおく．すると，次式が得られる (☆ 23)．

☆ 18) 記号 $\approx$ は，「近似的に等しい」という意味．

☆ 19) 最小二乗法: least squares method. 最小自乗法とも．

☆ 20) 二乗誤差: squared error. その平均は平均二乗誤差 (mean squared error, MSE).

☆ 21) 当てはまりの「良さ」の規準に二乗誤差を採用することにはちゃんとした理由があるが，この授業では省略する．

☆ 22) $\frac{1}{2}$ 倍してるのは，微分した後が楽になるように．

☆ 23) $E(a, b)$ は「下に凸な」関数なので，その偏導関数がいずれも 0 のときに最小となる．

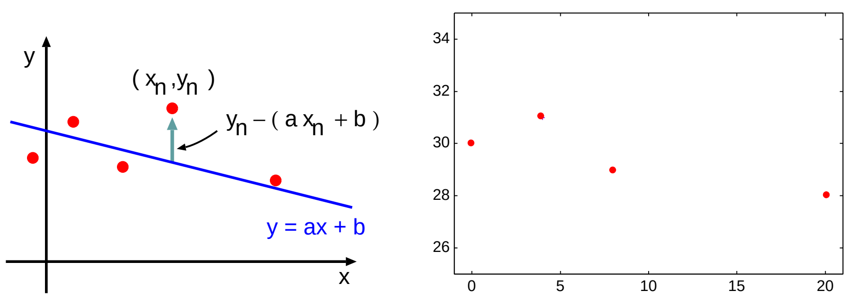


図 1: 左: 直線当てはめ，右: Q1 のデータを表した図

$$\frac{\partial E}{\partial a} = \sum_{n=1}^{N} (y_n - (ax_n + b))(-x_n) = a\sum_{n=1}^{N} x_n^2 + b\sum_{n=1}^{N} x_n - \sum_{n=1}^{N} x_n y_n = 0 \quad (6)$$

$$\frac{\partial E}{\partial b} = \sum_{n=1}^{N} (y_n - (ax_n + b))(-1) = a\sum_{n=1}^{N} x_n + b\sum_{n=1}^{N} 1 - \sum_{n=1}^{N} y_n = 0 \quad (7)$$

これを整理すると，次の連立一次方程式を得る．これを正規方程式という．

$$\begin{pmatrix} \sum_{n=1}^{N} x_n^2 & \sum_{n=1}^{N} x_n \\ \sum_{n=1}^{N} x_n & \sum_{n=1}^{N} 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{n=1}^{N} x_n y_n \\ \sum_{n=1}^{N} y_n \end{pmatrix} \quad (8)$$

これを解けば，学習データに対する二乗誤差を最小とするようなパラメータ a, b を得ることができる．

まとめると，この場合の最小二乗法の手続きは次のようになる．

1. 学習データ $\{(x_n, y_n) | n = 1, 2, \ldots, N\}$ を用いて正規方程式 (8) を求める．
2. その解 (a, b) を求める．

Q1. ほげお君は，こつこつためたお金で自動車を購入したいと思い，たまに中古車情報を調べている．ほげお君が気になっている自動車は，価格を調べ始めた日には 30 万円だったが，それ以降右の表のような価格推移を示している．調べ始めた日から経過した週数を x_n，価格差を y_n として，$n = 1, 2, 3, 4$ に対する (x_n, y_n) の値（例えば $(x_4, y_4) = (20, -2)$ である）に最小二乗法を適用してみよう．

経過週数	0	4	8	20
価格 [万円]	30	31	29	28
価格差 [万円]	0	1	-1	-2

(1) (x_n, y_n) $(n = 1, 2, 3, 4)$ の値から，直線のパラメータ a, b に対する正規方程式を求めなさい．

(2) (1) で求めた正規方程式を解いて a, b を求めなさい．

(3) (2) の結果を用いると，調べ始めから 36 週後には価格はいくらになっていると予測されるか答えなさい．

(5) (2) の結果を用いると，この自動車の価格が 25 万円になるのはいつ頃と予測されるか答えなさい．
