目次



前回の講義では，機械学習のアルゴリズムを「教師あり」「なし」の二つに大別した．教師あり学習では，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のに対して，教師なし学習では，学習データとして「入力」のみが与えられる．

# ★11　パターン認識と機械学習 (2) — 教師なし学習の例

教師なし学習の目的は，大量のデータが与えられたときに，そのデータがもつ規則性を見つけ出す，有益な情報を抽出する，といった処理を自動的に行うことである．その対象となる問題には様々なものがあるが，ここでは特に「クラスタリング」を取り上げる．

## ★11.1　$K$ 平均法によるクラスタリング

クラスタリングの手法にも様々なものがあるが，ここでは代表的なアルゴリズムとして，$K$ **平均法** (☆ 1) を紹介する．$K$ 平均法は，2 つのデータ $\boldsymbol{x}$ と $\boldsymbol{y}$ の間の非類似度 (☆ 2) を両者間のユークリッド距離 (☆ 3) $\|\boldsymbol{x}-\boldsymbol{y}\|$ で表せるようなデータに対して，それらを $K$ 個のクラスタに分ける分け方を見つけるための教師なし学習のアルゴリズムである．クラスタ数 $K$ はあらかじめ適当な方法で決めておかねばならない．$K$ 平均法の手順は次のようなものである．

1. 各学習データ $\boldsymbol{x}_n(n=1,2,\ldots,N)$ をランダムにクラスタ $C_1$ から $C_K$ までの $K$ 個のクラスタのいずれかに割り当てる．
2. クラスタ $C_k(k=1,2,\ldots,K)$ に割り当てられたデータの平均 $\boldsymbol{c}_k$ を求める．

$$\boldsymbol{c}_k = \frac{1}{|C_k|}\sum_{n:\boldsymbol{x}_n\in C_k}\boldsymbol{x}_n \tag{1}$$

   これをクラスタ $C_k$ の**セントロイド** (☆ 4) という．ここで，$|C_k|$ は集合 $C_k$ の要素数，すなわちクラスタ $C_k$ に割り当てられた学習データの数を表す．
3. 各学習データ $\boldsymbol{x}_n(n=1,2,\ldots,N)$ を，セントロイドとの距離が最小となるクラスタに割り当て直す．例えば $\boldsymbol{x}_n$ に対して

$$k^* = \underset{k=1,2,\ldots,K}{\operatorname{argmin}} \|\boldsymbol{x}_n-\boldsymbol{c}_k\|^2 \tag{2}$$

   であれば (☆ 5) (☆ 6)，$\boldsymbol{x}_n$ はクラスタ $C_{k^*}$ に割り当てる．
4. 上のステップの結果があらかじめ定めておいた条件を満たしている（後述）ならば終了，さもなくば 2. へ戻る．

この学習の終了条件としては，「クラスタ割り当てに変化がなくなった」，「ステップ 2,3 の実行回数が一定に達した」などが用いられる．また，$K$ 平均法では次式の $E$ の値 (これはクラスタリングの「誤差」を表している) が学習ステッ

☆ 1) $K$ 平均法: $K$-means 法とも．

☆ 2) 距離が小さい方が類似度が大きいので，「非」類似度が距離に対応すると考えている．

☆ 3) $\boldsymbol{x}$ が $D$ 次元ベクトル $\boldsymbol{x}=(x_1,x_2,\ldots,x_D)$ であり，$\boldsymbol{y}$ も同様の場合，$\|\boldsymbol{x}-\boldsymbol{y}\| = \sqrt{\sum_{d=1}^{D}(x_d-y_d)^2}$

☆ 4) セントロイド: centroid.

☆ 5) argmin の意味は次の例でわかるだろう．
$f(x)=(x-2)^2+3$ のとき，
$\min_x f(x)=3$,
$\operatorname{argmin}_x f(x)=2$.

☆ 6) ここでは距離の大小関係のみが問題なので，ユークリッド距離そのものではなく二乗した値で考えている（実際の計算では平方根が出てこない分その方が簡単だから）．

プ毎に単調減少することが知られているので，この値の減少幅が一定より小さくなったら終了する，という方法もよく用いられる．

$$E = \sum_{k=1}^{k} \sum_{n:\boldsymbol{x}_n \in C_k} \|\boldsymbol{x}_n - \boldsymbol{c}_k\|^2 \tag{3}$$

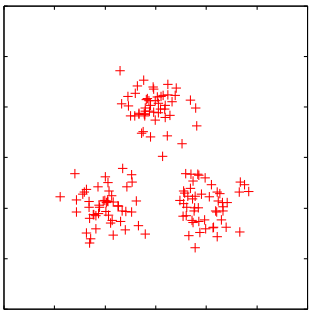
$K$ 平均法の結果は，学習データに対するクラスタ割り当ての初期値に依存する．そのため，実際には初期値を変えて何度か $K$ 平均法を実行し，上記の $E$ の値が最も小さかった結果を採用する，というような方法がとられる．

上記の学習手続きによってクラスタセントロイドが推定できたら，式 (2) と同様の計算によって未知データの所属クラスタも決めることができる．例えば，ある未知データ $\boldsymbol{x}$ について，

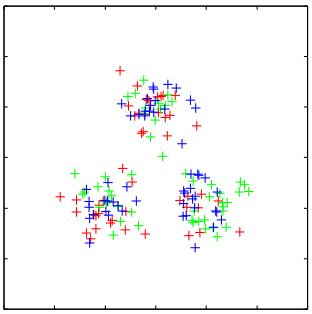
$$i = \underset{k=1,2,\ldots,K}{\operatorname{argmin}} \|\boldsymbol{x} - \boldsymbol{c}_k\|^2 \tag{4}$$

であれば，このデータの所属はクラスタ $C_i$ とすればよい．

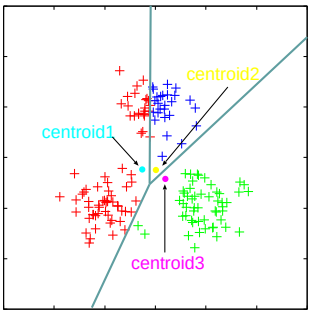
図 1 に，2 次元のデータに対して $K$ 平均法を適用した結果を示す．また，図 2 に，猫の顔画像 131 枚 (画素数は $64 \times 64$) に $K$ 平均法を適用して得られたセントロイドすなわちクラスタ毎の平均画像を示す．



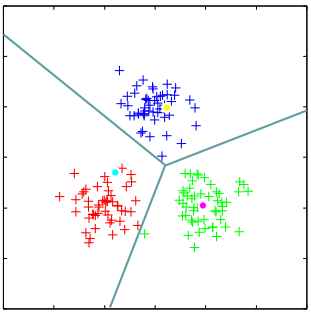
(1A) 学習データその 1



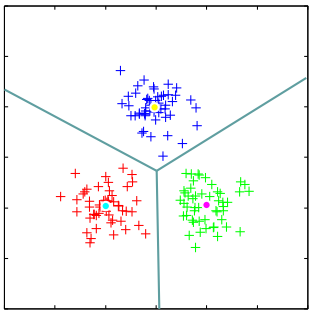
(1B) 初期値



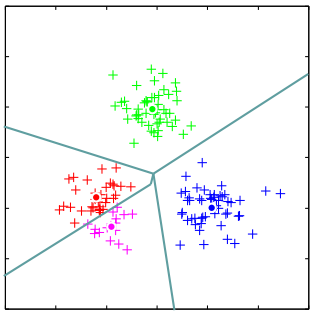

(1C) 学習 1 回



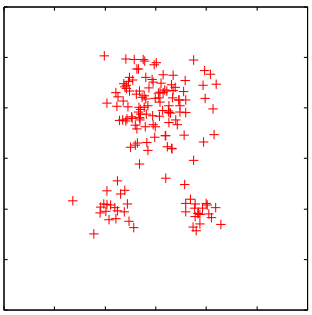
(1D) 学習 2 回



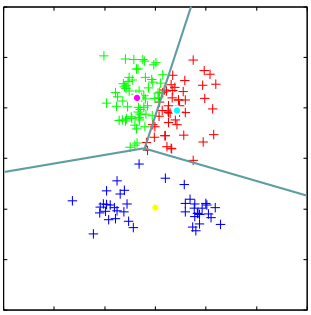
(1E) 学習 10 回



(2) $K = 4$ で学習 10 回



(3A) 学習データその 2



(3B) 学習 10 回

図 1: $K$ 平均法による 2 次元データのクラスタリング．(1A) は学習データの例．(1B) から (1E) までは，(1A) のデータ対して $K = 3$ で $K$ 平均法を適用した結果．(2) は，同じデータに $K = 4$ で $K$ 平均法を適用した結果．(3A) と (3B) は，別の学習データとそのクラスタリング結果 ($K = 3$)．

図 2: 131 枚の猫画像に $K$ 平均法を適用して得られたクラスタ平均画像 ($K = 5$)．

# ★12　パターン認識と機械学習 (3) — 回帰のための教師あり学習の例

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．今回は，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．

## ★12.1　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

$$y = f(x; a, b) = ax + b \tag{5}$$

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 3 参照)．つまり，$n = 1, 2, \ldots, N$ のそれぞれについて

$$y_n \approx ax_n + b \tag{6}$$

となるようにパラメータ $a$,$b$ を定める問題である (☆ 7)．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** (☆ 8) 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** (☆ 9) を考える (☆ 10)．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

$$(y_n - f(x_n))^2 = (y_n - (ax_n + b))^2 \tag{7}$$

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．(☆ 11)．

$$E(a, b) = \frac{1}{2} \sum_{n=1}^{N} (y_n - (ax_n + b))^2 \tag{8}$$

$E(a, b)$ 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 (☆ 12)．

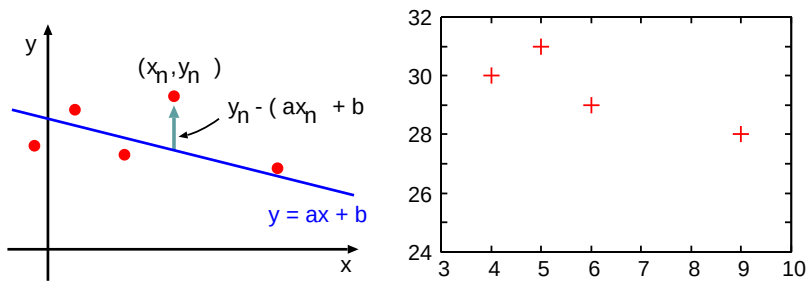



図 3: 左: 直線当てはめ，右: Q1 のデータを表した図

☆ 7) 記号 ≈ は，「近似的に等しい」という意味．

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

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

☆ 10) 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．

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

☆ 12) $E(a, b)$ 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．

$$\frac{\partial E}{\partial a} = \sum_{n=1}^{N}(y_n - (ax_n + b))(-x_n) = a\sum_{n=1}^{N} x_n^2 + b\sum_{n=1}^{N} x_n - \sum_{n=1}^{N} x_n y_n = 0 \quad (9)$$

$$\frac{\partial E}{\partial b} = \sum_{n=1}^{N}(y_n - (ax_n + b))(-1) = a\sum_{n=1}^{N} x_n + b\sum_{n=1}^{N} 1 - \sum_{n=1}^{N} y_n = 0 \quad (10)$$

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．

$$\begin{pmatrix} \sum_{n=1}^{N} x_n^2 & \sum_{n=1}^{N} x_n \\ \sum_{n=1}^{N} x_n & \sum_{n=1}^{N} 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{n=1}^{N} x_n y_n \\ \sum_{n=1}^{N} y_n \end{pmatrix} \quad (11)$$

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 (11) 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の価格推移は右のようになっていた．このデータに最小二乗法を適用してみよう．ただし，表の数をそのまま扱うと計算が面倒なので，次の手順に従ってみよう．

| 月 | 4 | 5 | 6 | 9 |
|---|---|---|---|---|
| 価格（万円） | 30 | 31 | 29 | 28 |

(1) 「4 月から何ヶ月後か」を $x_n$，「4 月の価格との差 (単位は万円)」を $y_n$ として，$n = 1, 2, 3, 4$ に対する $x_n, y_n$ の値の表を書きなさい (例えば $(x_4, y_4) = (5, -2)$ である)．

(2) (1) 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．

(3) (2) で求めた正規方程式を解いて $a, b$ を求めなさい．

(4) (3) の結果を用いると，来年 3 月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) (3) の結果を用いると，この車の価格が 20 万円になるのはいつ頃と予測されるか答えなさい．

---

## ★ 宿題

**Q2.** ほげお君は，長年の研究の結果，人間の性格を

$$(\text{「ほげ度」},\text{「ふが度」},\text{「へな度」}) \tag{12}$$

という 3 つの数値をならべた 3 次元ベクトルで特徴づけられることを発見した．彼は，たくさんの人のこれらの数値を測定し，得られた大量の 3 次元ベクトルから成るそのデータに $K$ 平均法を適用して，次の 3 つのセントロイド $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$ を得た（すなわち $K = 3$ である）．

$$\boldsymbol{c}_1 = (1, 2, 3) \qquad \boldsymbol{c}_2 = (0, -2, 0.5) \qquad \boldsymbol{c}_3 = (-1, 0, -2) \tag{13}$$

これらのセントロイドは，それぞれのクラスタを代表する値とみなせる．このとき，次の人物はどのセントロイドが代表するクラスタに所属するといえるか．

ほげほげお: $(\text{「ほげ度」},\text{「ふが度」},\text{「へな度」}) = (0, 0, 0)$
たかたか: $(\text{「ほげ度」},\text{「ふが度」},\text{「へな度」}) = (1, 1, 5)$

また，上記の 2 人とは異なるクラスタに属する人物を表す 3 次元ベクトルを適当に定め，その人物がどのクラスタに属するかを計算して示しなさい．

**Q3.** 式 (5) のかわりに 3 つのパラメータ $a, b, c$ をもつ放物線

$$y = f(x; a, b, c) = ax^2 + bx + c \tag{14}$$

を考えれば，最小二乗法による放物線の当てはめができる．この場合，学習データ $\{(x_n, y_n) | n = 1, 2, \ldots, N\}$ に対する正規方程式は次式で与えられる（ここでは $\sum$ は省略した書き方をしている）ことを示しなさい．

$$\begin{pmatrix} \sum x_n^4 & \sum x_n^3 & \sum x_n^2 \\ \sum x_n^3 & \sum x_n^2 & \sum x_n \\ \sum x_n^2 & \sum x_n & \sum 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum x_n^2 y_n \\ \sum x_n y_n \\ \sum y_n \end{pmatrix} \tag{15}$$

ヒント: 誤差関数を定義して，偏微分して 0 とおいて…