目次

★ 14　パターン認識と機械学習 (5) — 教師なし学習（承前）

★ 14.3　主成分分析と次元圧縮

次元圧縮の手法にも様々なものがあるが，ここでは，統計的なデータ分析手法である**主成分分析** (☆ 1) に基づく手法を紹介する．主成分分析の目的は，前回資料式 (10) のデータから複数の互いに「無関係」な因子を取り出し，個々のデータをこれらの因子の線形結合で表すことである．

☆ 1) 主成分分析: Principal Component Analysis.

簡単のため，以下では因子を一つだけ取り出す場合について説明する．この場合，主成分分析の問題設定は次のようになる：「前回資料式 (10) のデータ $\boldsymbol{x}_n$ をベクトル $\boldsymbol{a}$ を用いて次式のようにスカラ y_n に変換するとき，$\{y_n\}$ の分散が最大となるように $\boldsymbol{a}$ を定めよ．ただし，$\bar{\boldsymbol{x}}$ はデータの平均すなわち $\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n$ である．」

$$y_n = \boldsymbol{a}^\top(\boldsymbol{x}_n - \bar{\boldsymbol{x}}) \tag{1}$$

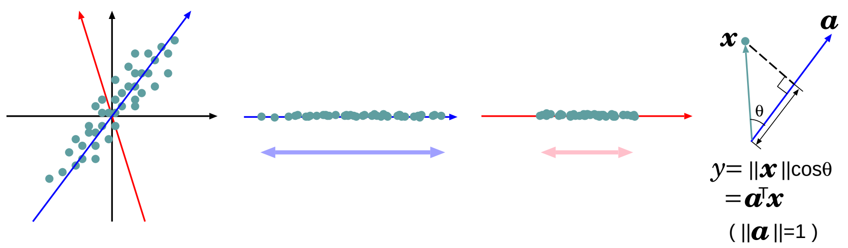

図 1: 左端の図に灰色で示された 2 次元データを青い軸方向の成分のみで説明しようとすると，その成分は左から 2 番目の図のように散らばる．赤い軸方向の成分のみだと，左から 3 番目の図のように散らばる．

この問題の解 $\boldsymbol{a}$ がどのようなベクトルになるか考えよう．まず，式 (1) より，y_n の平均は 0 である (☆ 2)．したがって，$\{y_n\}$ の分散は $\frac{1}{N}\sum_{n=1}^{N} y_n^2$ と書け，

☆ 2) $\sum_{n=1}^{N} y_n$
$= \sum_{n=1}^{N}\boldsymbol{a}^\top\boldsymbol{x}_n - \sum_{n=1}^{N}\boldsymbol{a}^\top\bar{\boldsymbol{x}}$
$= \boldsymbol{a}^\top\sum_n\boldsymbol{x}_n - \boldsymbol{a}^\top\sum_n\bar{\boldsymbol{x}}$
$= N\boldsymbol{a}^\top\bar{\boldsymbol{x}} - N\boldsymbol{a}^\top\bar{\boldsymbol{x}} = 0$

$$\frac{1}{N}\sum_{n=1}^{N} y_n^2 = \frac{1}{N}\sum_{n=1}^{N}\left(\boldsymbol{a}^\top(\boldsymbol{x}_n - \bar{\boldsymbol{x}})\right)\left(\boldsymbol{a}^\top(\boldsymbol{x}_n - \bar{\boldsymbol{x}})\right)^\top \tag{2}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{a}^\top(\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^\top\boldsymbol{a} \tag{3}$$

$$= \boldsymbol{a}^\top\left(\frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^\top\right)\boldsymbol{a} = \boldsymbol{a}^\top\boldsymbol{V}\boldsymbol{a} \tag{4}$$

と変形できる（$\boldsymbol{V}$ はデータの**分散共分散行列**である）．すなわち，求めるべきは，$\boldsymbol{a}^\top\boldsymbol{V}\boldsymbol{a}$ を最大にするベクトル $\boldsymbol{a}$ である．ただし，この値は $\boldsymbol{a}$ の長さを大きくすればいくらでも大きくできるので，$\|\boldsymbol{a}\|$ に条件を付けないと意味がない．そこで，$\|\boldsymbol{a}\| = 1$ という制約条件のもとで $\boldsymbol{a}^\top\boldsymbol{V}\boldsymbol{a}$ を最大化する $\boldsymbol{a}$ を求めよう．

上記のような制約条件付きの最適化問題を解く定番の手法は，ラグランジュの未定乗数法である．この問題の場合，ラグランジュ乗数を λ として

$$L = \boldsymbol{a}^\top \boldsymbol{V}\boldsymbol{a} - \lambda(\boldsymbol{a}^\top \boldsymbol{a} - 1) \tag{5}$$

とおけば，$\frac{\partial L}{\partial \boldsymbol{a}} = \boldsymbol{0}$ を満たす $\boldsymbol{a}$ が解の候補となる．ここで，

$$\frac{\partial L}{\partial \boldsymbol{a}} = 2\boldsymbol{V}\boldsymbol{a} - 2\lambda\boldsymbol{a} \tag{6}$$

なので (☆ 3)，解は

$$\boldsymbol{V}\boldsymbol{a} = \lambda\boldsymbol{a} \tag{7}$$

を満たさねばならない．すなわち，解候補は行列 $\boldsymbol{V}$ の単位固有ベクトルである．さらに，式 (7) を式 (5) に代入すると

$$L = \lambda\boldsymbol{a}^\top\boldsymbol{a} - \lambda\cdot 0 = \lambda \tag{8}$$

となることから，この問題の解は，$\boldsymbol{V}$ の最大固有値に対する単位固有ベクトルであることがわかる (☆ 4)．

ここではこれ以上深入りしないが，複数の因子を取り出す場合についても同様の議論ができる．その場合，$\boldsymbol{V}$ の固有ベクトルを対応する固有値の大きい方から順に選んで，それらを因子として採用すればよいことがわかる．

☆ 3) $\frac{\partial L}{\partial \boldsymbol{a}}$ は,「L を $\boldsymbol{a}$ の各要素で微分したものをならべたベクトル」である．要素ごとの微分を考えると，$\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x}^\top\boldsymbol{x} = 2\boldsymbol{x}$ 等の公式を導出できる．

☆ 4) $\boldsymbol{V}$ は半正値対称行列なので固有値は必ず 0 以上の実数になる（詳細は省略）．

以下に示すのは，猫の顔画像に主成分分析を適用した例である（第 4 回講義資料と同じもの）．ここで「固有顔」と呼んでいるものは，猫の顔画像の学習データから求めた分散共分散行列の固有ベクトルである．

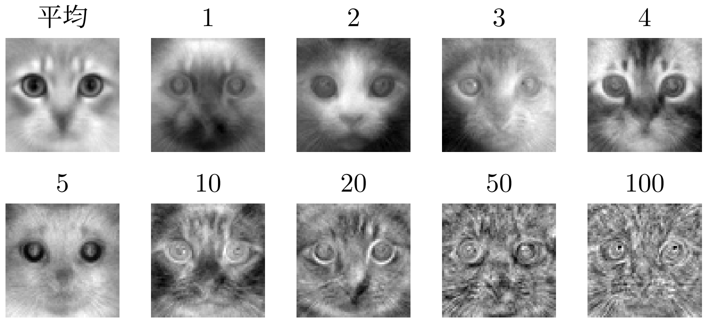

図 1: 猫の固有顔の例．左上の画像は多数の猫の顔画像の平均であり，$1, 2, \ldots$ と記された画像は固有顔を可視化したもの（番号の小さいものほど対応する固有値が大きく,「重要」な基底）．これらの固有顔は正規直交基底を成している．

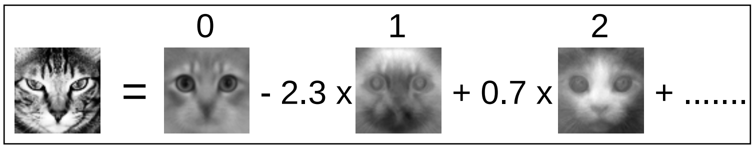

図 2: 固有顔を用いた猫の顔の展開の概念図．画像の上の番号は，固有顔の番号に対応している（0 と記された画像は平均）．

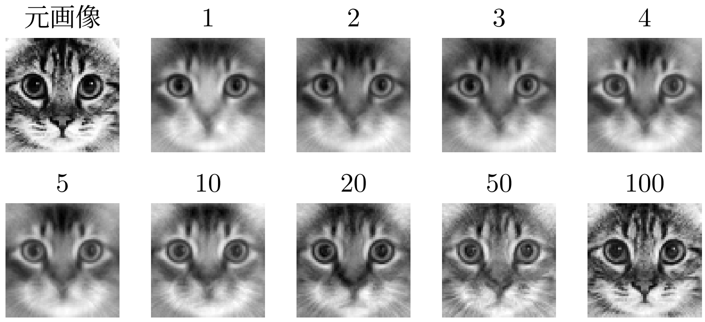

図 3: 固有顔を用いた近似．左上は 4096 個の画素値から成る元画像．それ以外は，画像の上に記された数の基底のみを用いて元画像を近似したもの．例えば，5 と記された画像は，上記の 1 番から 5 番までの基底のみを用いて元画像を近似しており，5 つの基底に対応する 5 つの展開係数の値のみでこの画像の情報を表現していることになる．

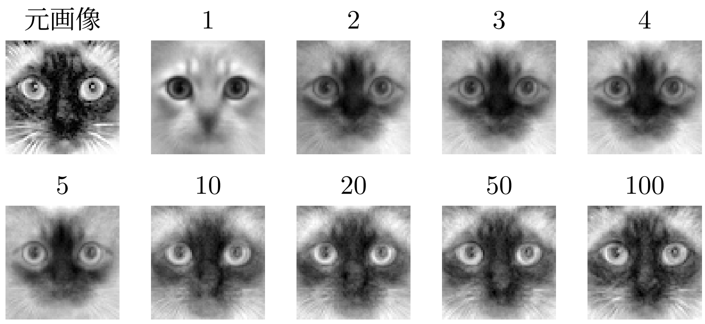

図 4: もうひとつの例．