Algorithmic Marketing P44-45

Hiroo Takizawa

1. p43下5行からp44 12行目

- 共分散行列について
- 共分散行列が対角行列でなければ、各特徴量は相関している
- やりたいことは、Xを変換し、共分散行列が対角行列のような、n x mの計画行列Zを作ること。
- 主成分分析のアプローチでは線形変換によって得られると思われる。

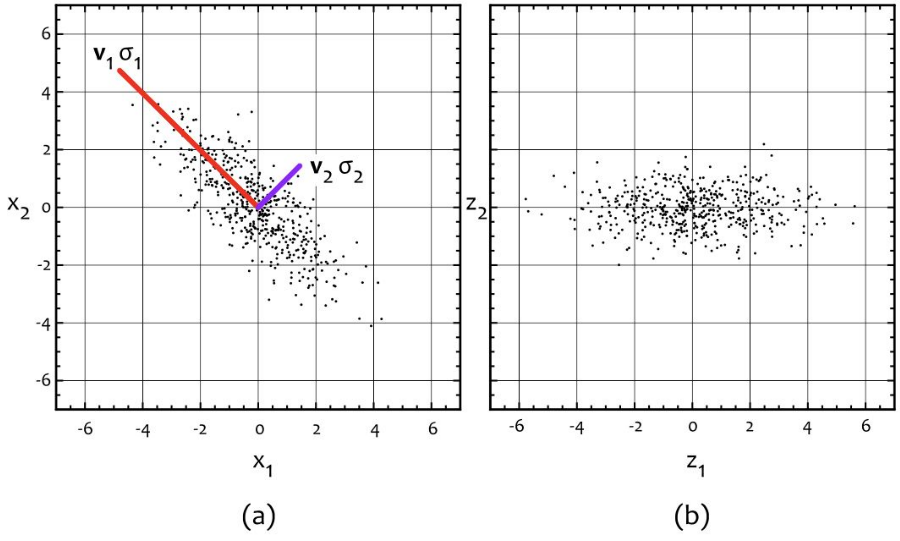
$$\mathbf{Z} = \mathbf{XT} \tag{2.64}$$

- Tは変換行列。
- Tを作るにははじめに、計画行列がどのように計画行列におけるムンさんが最大となるような方向に対応するベクトル基底に因数分解されるかをを見る。

2. p44 13行目ーp44 下から8行目

- はじめに、データの分散が最大となる方向を見つける。
- 散布図におけるメインの軸。
- まず、下記の条件を満たすm次元の単位ベクトルを見つける。これは分散が最大の軸。
-

$$\mathbf{v}_1 = \underset{\mathbf{v}}{\operatorname{argmax}} \ \|\mathbf{X}\mathbf{v}\|^2, \qquad \|\mathbf{v}\| = 1$$



2. p44 13行目ーp44 下から8行目

- 次に、v1ベクトルに直行する2つ目の単位ベクトルを考える。残った分散を得るために。

$$\mathbf{v}_2 = \underset{\mathbf{v}}{\operatorname{argmax}} \ \|\mathbf{X}\mathbf{v}\|^2, \qquad \|\mathbf{v}\| = 1 \ \text{ and } \ \mathbf{v}\cdot\mathbf{v}_1 = 0 \tag{2.66}$$

- このプロセスを何度も行う。それまでに得たベクトル全てに直行するようなベクトルを考える。
- 計画行列の列数が r なら、r個のゼロでないベクトルが得られる
- これらのベクトルはそれぞれ、計画行列に残った分散が最大となるような軸に対応している。

3. p44 下から7行目-p45下から11行目

- ここまでで得た行ベクトルを結合し、m x rの行列とする。これをVとする。
- これらの単位ベクトルはそれぞれ直行するので、

$$\mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I} \tag{2.67}$$

- 単位ベクトルは向きは示せるものの、大きさは表現できない。大きさは下記で求める。

$$\sigma_i = \|\mathbf{X}\mathbf{v}_i\| \tag{2.68}$$

3. p44 下から7行目-p45下から11行目

- 主成分は最初が最も大きく、徐々に小さくなる。

$$\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r \tag{2.69}$$

- 対角成分にこれを置いた r x r の対角行列をΣとする。

$$\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r) \tag{2.70}$$

4. p45下から10行目-p45最後

- ここまでで、ベクトルVの正規直交基底と、それに対応するスケール因子Σを得た。
- 計画行列の因数分解を完成させるための3つ目の因子が必要。それは、計画行列を主成分基底に写像する、あるいは、基底を混ぜて計画行列にするような3つ目の因子が必要である。
- この因子をUとすると、

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}} \tag{2.71}$$

- Uについてこれを解くと

$$\mathbf{U} = \mathbf{X}\mathbf{V}\boldsymbol{\Sigma}^{-1} \tag{2.72}$$