Stat 315c: Transposable Data
Principal Components (review)

Art B. Owen

Stanford Statistics

Principal Components Analysis

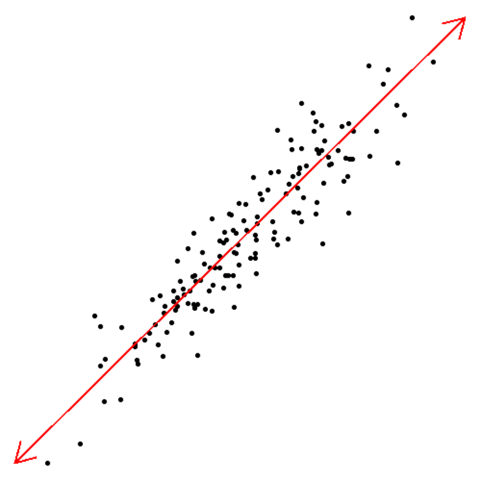
- Principal components are a core technique for dimension reduction
- They find a space of reduced dimension that still captures most of the variance of the data
- We'll use it at first to understand some classical methods.
- Then we revisit it as an 'end in itself'

P.C. Illustration

Some points in $\mathbb{R}^2$

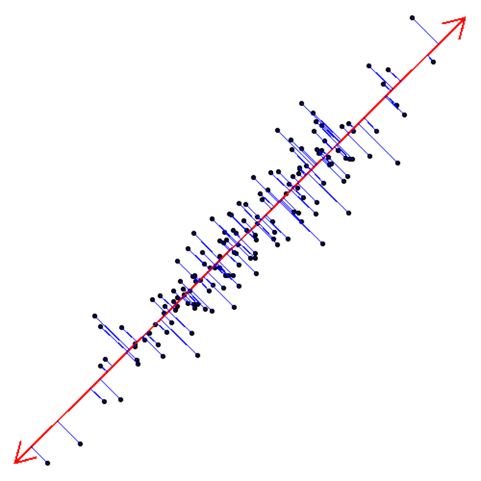
P.C. Illustration

They vary largely in one direction



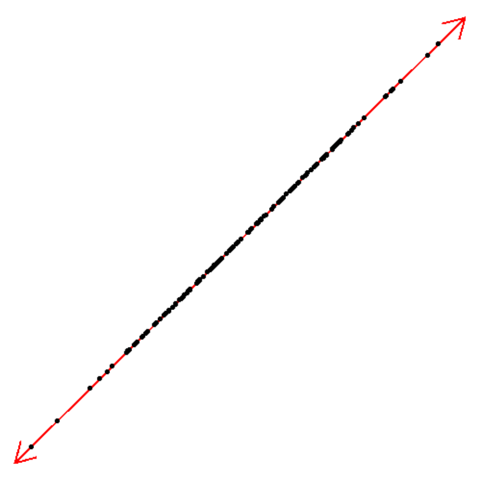
P.C. Illustration

We project orthogonally to that direction



P.C. Illustration

Finding a one dimensional summary



More generally

Recipe

- We have n points in $\mathbb{R}^d$
- Put them in $X_{n\times d}$ one point per row
- Project the points onto k dimensional subspace of $\mathbb{R}^d$ spanned by $V_{1:k}$
- $X' \to V_{1:k}V_{1:k}'X'$ n points in $\mathbb{R}^d$ $V_{1:k}$ has k orthonormal columnns
- So $X \to XV_{1:k}V_{1:k}' = U\Sigma V'V_{1:k}V_{1:k}' = U_{1:k}\Sigma_{1:k}V_{1:k}'$
- Throw away $d-k$ dimensions
- Just keep an $n\times k$ matrix sometimes $k \ll d$

Example

- Measure light energy in 200 bands
- Find most of the natural variation is only in 3 dimensions

PC dimension reduction

Uses

- Interpret flavors of variation
- Regress Y on principal components of X only
- Reduce to principal components, then cluster
- Neither saves measurement costs
- Either could be more efficient
- Or miss the point entirely

Population Principal components

First principal component

- Maximize variance of $u'x$ over $u \in \mathbb{R}^d$ with $u'u = 1$
- For $\mathrm{Var}(x) = \Sigma$, find

$$\max_{u'u=1} u'\Sigma u$$

Eigendecomp of Σ

- Recall $\frac{u'\Sigma u}{u'u}$ maximized for $u =$ first eigen vector
- Write $\Sigma = P\Lambda P'$ with $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_d)$ and $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_d \ge 0$

$$P = \begin{pmatrix} u_1 & u_2 & \cdots & u_d \end{pmatrix}$$

- First principal component is u_1 with $u_1'\Sigma u_1 = \lambda_1$
- Begin proof: Let $u = \sum_j \alpha_j u_j$ with $\sum_j \alpha_j^2 = 1 \cdots$

Further principal components

Second P.C.

- Maximize $u'\Sigma u$ subject to $u'\Sigma u_1 = 0$
- That way $x'u$ is uncorrelated with $x'u_1$
- Get $u = u_2$, second eigenvector
- More generally: k'th P.C. vector is u_k for $k = 1, \dots, d$. It max's $u'\Sigma u$ st $u'\Sigma u_j = 0$ for $1 \le j < k$

Variances: $x = (x_1, \dots, x_d) \in \mathbb{R}^d$

- $\mathrm{Var}(x'u_k) = u_k'\Sigma u_k = u_k'(\lambda_k u_k) = \lambda_k$
- $\sum_{j=1}^d \mathrm{Var}(x_j) = \mathrm{tr}(\Sigma) = \sum_{j=1}^d \lambda_j = \sum_{j=1}^d \mathrm{Var}(u_j'x)$
- $$\text{Keep } k \text{ P.C.s} \implies \text{keep } \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^d \lambda_j} \text{ of the variance}$$

Sample Principal components

Just replace Σ by

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'$$

Scale issue

- Change x_1 from meters to femtometers
- Scales x_1 by 10^{15}
- Increases Var(x_1) by 10^{30}
- Could changes principal components in an important way

Can apply P.C. to any variance matrix

Sometimes use

- Correlation matrix

$$\hat{R} = \mathrm{diag}(\hat{\Sigma})^{-1/2}\,\hat{\Sigma}\,\mathrm{diag}(\hat{\Sigma})^{-1/2}$$

- Outer product matrix

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i'$$

- Variance within groups

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}(i))(x_i - \hat{\mu}(i))'$$

- Variance between groups
- Residual variance matrix

Computing P.C. from SVD

$$x_i \in \mathbb{R}^d, \quad i = 1, \dots, n$$

$$\widetilde{X} = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}$$

$$\widetilde{X} = U\Sigma V' \quad \text{SVD}$$

$$\hat{\Sigma} = \frac{1}{n}\widetilde{X}'\widetilde{X} = \frac{1}{n}V\Sigma'\Sigma V'$$

Take cols of V and $\dfrac{\sigma_j^2}{n}$