$\boxed{\text{PCA}}$ n data points, p features

original data $\quad \underline{x}_1 = \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{pmatrix} \quad \underline{x}_2 = \begin{pmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2p} \end{pmatrix} \quad \cdots \quad \underline{x}_n = \begin{pmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{np} \end{pmatrix}$

Assume data centered, $\underline{x}_i \leftarrow \underline{x}_i - \overline{\underline{x}}$

Let $\quad X = \begin{bmatrix} \underline{x}_1^T \\ \vdots \\ \underline{x}_n^T \end{bmatrix} \quad n \times p$

$$\underline{x}_i = x_{i1}\,\underline{e}_1 + x_{i2}\,\underline{e}_2 + \cdots + x_{ip}\,\underline{e}_p$$

$\vec{e}_1, \vec{e}_2, \dots, \vec{e}_p \to$ old, or original features

$$= \underbrace{z_{i1}}\,\underbrace{\underline{a}_1} + \underbrace{z_{i2}}\,\underbrace{\underline{a}_2} + \cdots + \underbrace{z_{ip}}\,\underbrace{\underline{a}_p}$$

$\vec{a}_1, \vec{a}_2, \dots, \vec{a}_p \to$ new ~~features~~ features

$\underline{a}_1, \underline{a}_2, \dots, \underline{a}_p \to$ eigenvectors

$z_{i1}, z_{i2}, \dots, z_{ip}$: Principal components or scores or component scores

$\{z_{ij}\}$ and $\{\underline{a}_j\}$ come from

$$\hat{\Sigma} = \frac{1}{n-1} X^T X = A^T D A \quad \text{(eigendecomposition)}$$

with $A = \begin{bmatrix} \underline{a}_1^T \\ \underline{a}_2^T \\ \vdots \\ \underline{a}_p^T \end{bmatrix}$ eigenvectors and $\underline{z}_i = A \underline{x}_i$

and $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$

eigenvalues

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$

new feature $\qquad$ new feature $\qquad$ new feature

$$
\begin{aligned}
\underline{x}_1 &= z_{11}\,\underline{a}_1 + z_{12}\,\underline{a}_2 + \cdots + z_{1p}\,\underline{a}_p \\
\underline{x}_2 &= z_{21}\,\underline{a}_1 + z_{22}\,\underline{a}_2 + \cdots + z_{2p}\,\underline{a}_p \\
&\ \vdots \\
\underline{x}_n &= z_{n1}\,\underline{a}_1 + z_{n2}\,\underline{a}_2 + \cdots + z_{np}\,\underline{a}_p
\end{aligned}
$$

variance $\frac{1}{n-1}\sum_{i=1}^{n} z_{i1}^2 = \lambda_1$ $\qquad$ variance λ_2 $\qquad \cdots \qquad$ variance λ_p

Variances are sorted, so are the scores $\Rightarrow$ 1st scores tend to be largest, 2nd scores large, but slightly smaller, ..., last scores are small.

Variance decomposition

Note the total variance of $\underline{x}_1, \ldots, \underline{x}_n$ is

(sum of variances of each feature)

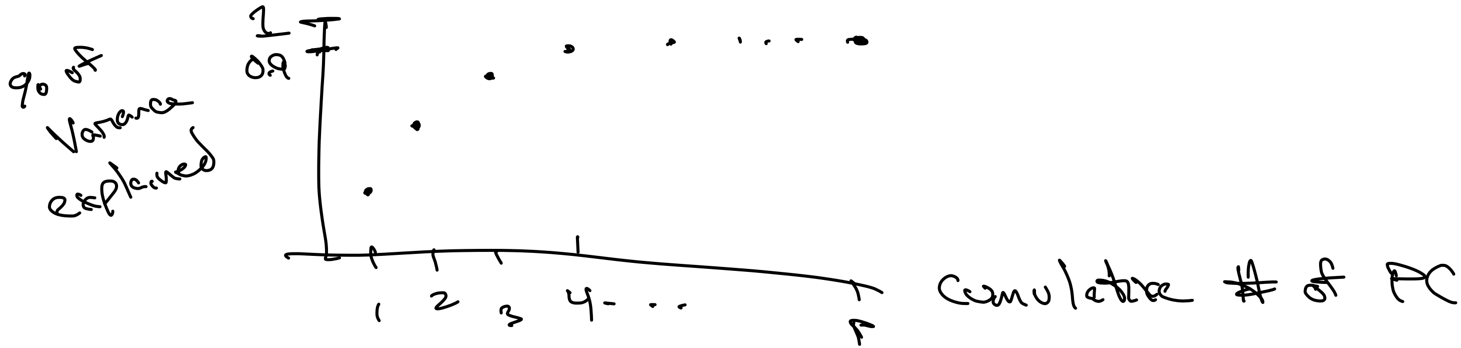
$$= \frac{1}{n-1}\sum_{i=1}^{n} x_{i1}^2 + \frac{1}{n-1}\sum_{i=1}^{n} x_{i2}^2 + \cdots + \frac{1}{n-1}\sum_{i=1}^{n} x_{ip}^2$$ [b/c features centered!]

$$= tr(\hat{\Sigma}) = tr(A^T D A) = tr(D A A^T) = tr(D)$$

$$= \lambda_1 + \lambda_2 + \cdots + \lambda_p$$

Percent of variance explained by 1st k principal components:

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_k + \lambda_{k+1} + \cdots + \lambda_p}$$



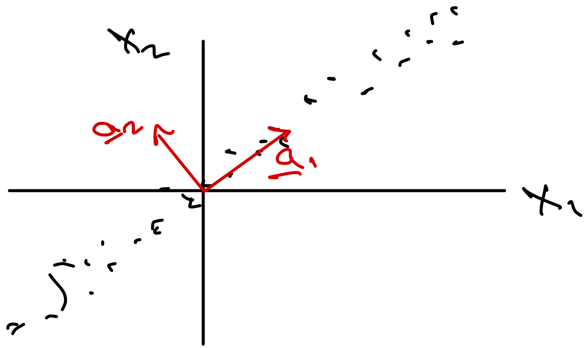
If % of variance explained by 1st k PCs is big

[80%, 90%, e.g.] , then

$$\underline{x}_i = z_{i1}\underline{a}_1 + z_{i2}\underline{a}_2 + \cdots + z_{ik}\underline{a}_k + \cdots + z_{ip}\underline{a}_p$$

$$\simeq z_{i1}\underline{a}_1 + z_{i2}\underline{a}_2 + \cdots + z_{ik}\underline{a}_k$$

discard



λ_1 big ≈ 10

λ_2 small ≈ 0.7

% of var. PC1 explains: $\dfrac{10}{10+0.7} \approx 0.93$

$\underline{a}_1, \underline{a}_2, \dots, \underline{a}_p$ give directions of decreasing variability

$\lambda_1 \underline{a}_1, \lambda_2 \underline{a}_2, \dots, \lambda_p \underline{a}_p$ = "loadings"

Principal Component Regression

Original regression problem $\quad Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon$

Do a PCA decomposition of features, if see

$k \ll p$ explains "most" of the variance in data,

??

then

$$Y = \beta_0 + \beta_1 z_1 + \dots + \beta_k z_k + \varepsilon$$

Rules of thumb

- Use enough PCs to explain 80% or 90% of variation
- Retain components whose eigenval λ_i are greater than avg var $\frac{1}{p}\sum_{i=1}^{p}\lambda_i$
- Scree graph (% of var expl vs. component #) look for natural breaks.
- Test for significance of "large" components.