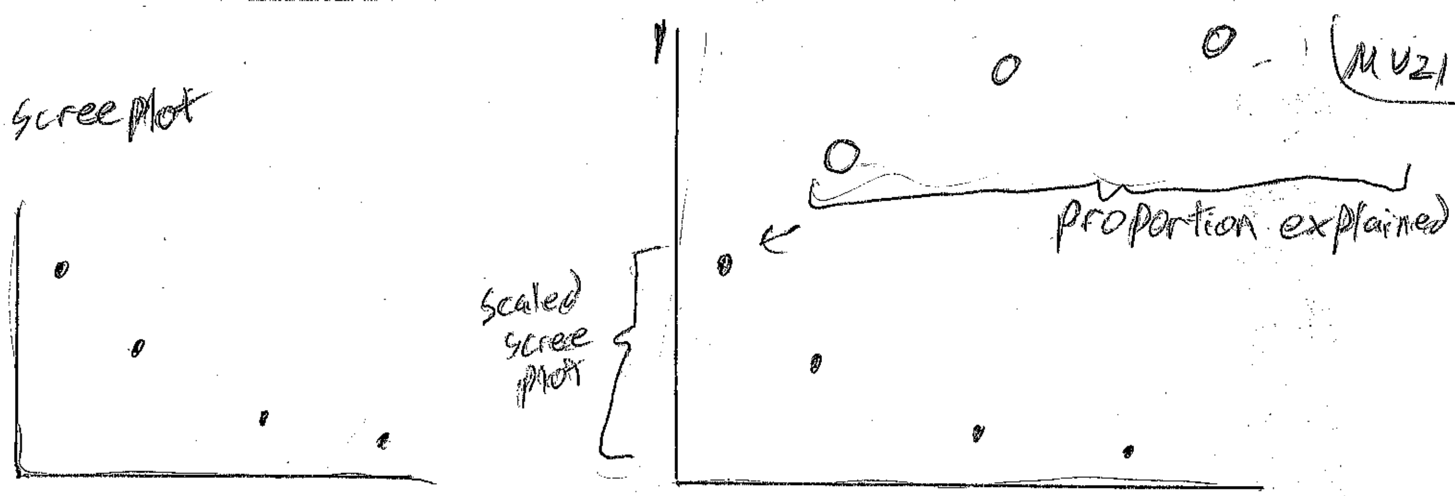
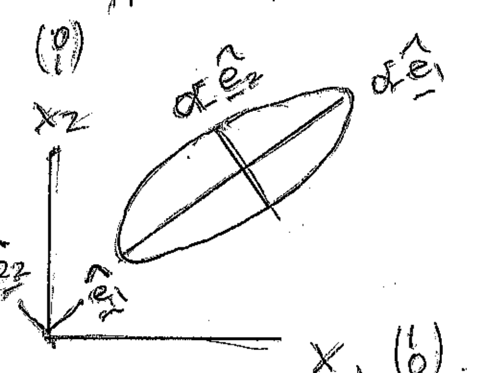


8) p126 Let $\hat{\Sigma}$ have eigenvalue eigenvector pairs $(\hat{\lambda}_1, \hat{e}_1), \ldots, (\hat{\lambda}_p, \hat{e}_p)$. The p <u>principal components</u> corresponding to the jth case $\underline{x}_j$ are $z_{j1} = \hat{\underline{e}}_1^T \underline{x}_j, \ldots, z_{jp} = \hat{\underline{e}}_p^T \underline{x}_j$.
The jth principal component coefficient vector is $\hat{\underline{e}}_j$ = pcj.
The <u>population principal components</u> corresponding to Σ are $Y_{ji} = \underline{e}_i^T x_j$. so $z_{ji} = \hat{Y}_{ji}$.

9) Let $\underline{z}_j = (z_{j1}, \ldots, z_{jp})^T$. New data matrix $Z = \begin{pmatrix} \underline{z}_1^T \\ \vdots \\ \underline{z}_n^T \end{pmatrix}$ corresponds to a rotation of the coordinate axes to $\hat{\underline{e}}_1, \ldots, \hat{e}_p$, the axes of the hyperellipsoid $\{\underline{z} \mid (\underline{z}-\underline{T})^T \hat{\Sigma}^{-1} (\underline{z}-\underline{T}) \le h^2\}$.
often take $T = \underline{0}$.



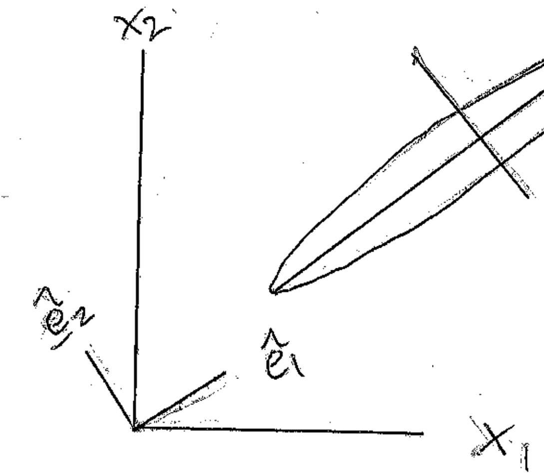
10) The principal components $z_1, \ldots, z_p$ act as p new variables created from $x_1, \ldots, x_p$.

11) Consider the covering hyperellipsoid $\{\underline{z} \mid D_{\underline{z}}(T, C) \le D_{(n)}(T, C)\}$.
Let $T = 0$ or consider centered data $\underline{w}_i = \underline{x}_i - T$.

2/15

The ith principal component is the projection of the data on the ith axis of the hyperellipsoid.

want <u>dimension reduction</u>: use $k < p$ principal components.



12] p128 Suppose you need to approximate the p variables by 1 variable. The projection on the major axis is "best" and has maximum variability. For 2 variables, project on major axis and next largest axis. It can be shown that if S or R is used, the 1st principal component $\hat{\underline{e}}_1^T \underline{x}$ is the linear combination $\underline{g}_1^T \underline{x}$ that maximizes $Var(\underline{g}_1^T \underline{x})$ subject to $\underline{g}_1^T \underline{g}_1 = 1$, while the jth principal component $\hat{\underline{e}}_j^T \underline{x}$ is the linear combination $\underline{g}_j^T \underline{x}$ that maximizes $Var(\underline{g}_j^T \underline{x})$ subject to $\underline{g}_j^T \underline{g}_j = 1$ and $cov(\underline{g}_j^T \underline{x}, \underline{g}_k^T \underline{x}) = 0$ for $k < j$.

$p \leq 11$

13] So a plot of PC1 vs PC2 is a good 2 dim summary plot. A scatterplot matrix of $\hat{\underline{e}}_1^T \underline{x}, \ldots, \hat{\underline{e}}_p^T \underline{x}$ is useful.

14) Let U be the subset of the data from which the robust estimator (T, C) is computed. Let S_U and R_U denote the sample covariance matrix and sample correlation matrix computed from the cases in U. Then $C = d\, S_U$ for some constant $d > 0$ and R_U is the generalized correlation matrix corresponding to C. The robust PCA uses U corresponding to RMVN.

15) p130 If $\mathrm{cov}(\underline{x}) = \Sigma_{\underline{x}}$, then on a large class of EC distributions,

$$S_U \xrightarrow{D} a\Sigma_{\underline{x}} \quad \text{and} \quad R_U \xrightarrow{D} \rho_{\underline{x}}.$$

Hence $\hat{\lambda}_i(R) - \hat{\lambda}_i(R_U) \xrightarrow{D} 0.$

Since $a\Sigma_{\underline{x}}\, \underline{e} = (a\lambda)\underline{e}$, $\frac{1}{a}\hat{\lambda}_i(S_U) - \hat{\lambda}_i(S) \xrightarrow{D} 0.$

So $\mathrm{corr}\left[(\hat{\lambda}_1, \ldots, \hat{\lambda}_p), (\hat{\lambda}_1, \ldots, \hat{\lambda}_p)_{Rob}\right] \xrightarrow{P} 1.$

16) p125 Let $p \times p\ \Sigma > 0$.

a) If $\hat{\Sigma} \xrightarrow{P} \Sigma$, then $\hat{\Sigma}\underline{e}_i - \hat{\lambda}_i \underline{e}_i \xrightarrow{P} 0$

b) If $\hat{\Sigma} \xrightarrow{P} \Sigma$, then $\Sigma\hat{\underline{e}}_i - \lambda_i \hat{\underline{e}}_i \xrightarrow{P} 0$

If $\hat{\Sigma} - \Sigma = O_P(n^{-\frac{1}{2}})$, then c) $\lambda_i \underline{e}_i - \hat{\Sigma}\underline{e}_i = O_P(n^{-\frac{1}{2}})$ and

d) $\hat{\lambda}_i \hat{\underline{e}}_i - \Sigma\hat{\underline{e}}_i = O_P(n^{-1/2})$.

22.5

e) If $\hat{\Sigma} \xrightarrow{P} c\Sigma$ for some constant $c > 0$ and if $\lambda_1 > \cdots > \lambda_p > 0$ are unique, then

$$|\text{corr}(\hat{e}_j, e_j)| \xrightarrow{P} 1.$$

So asymptotically the eigenvalues and eigenvectors of $\hat{\Sigma}$ act like those of Σ and vice verca. This result is useful since eigenvectors are not continuous functions of C. Scree plots and proportion of trace explained should be similar for robust and classical PCA.

17)

R or R_U	S or S_U
$tr(\hat{\Sigma}) = p$	$tr(\hat{\Sigma}) = \sum_{i=1}^p \hat{\lambda}_i$
$Y_i = \hat{e}_i^T x$ $\text{Corr}(Y_i, X_k) = \hat{e}_{ki}\sqrt{\hat{\lambda}_i} \propto \hat{e}_{ki}$ ith PC, kth random variable	$\text{Corr}(Y_i, X_k) = \dfrac{e_{ki}\sqrt{\hat{\lambda}_i}}{\sqrt{\hat{\sigma}_{kk}}}$
useful if the $\hat{\sigma}_{kk}$ differ (or are similar)	most useful if the $\hat{\sigma}_{kk}$ are similar

ex) p127 $\text{cov}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ so $\dfrac{\lambda_i}{\lambda_1 + \lambda_2} = \dfrac{1}{2}$.

If X_j is multiplied by 3, then $V(X_j) = 9$ and

$\dfrac{\lambda_j}{\lambda_1 + \lambda_2} = 0.9$, falsely suggesting X_j is more important.

$\lambda_1 + \lambda_2 = 1 + 9 = 10$ $\quad \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 9\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ so $\lambda_2 = 9$

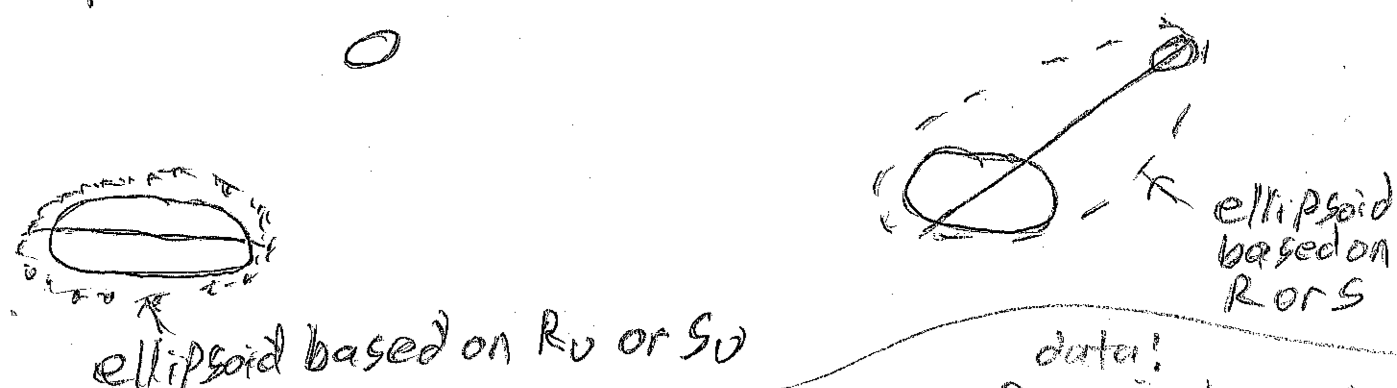
18) p129 If $\overbrace{\underline{w}_j = \underset{\sim}{x}_j - \underline{T}}^{\text{centering}}$ for $T = \underline{\bar{x}}$ or $\bar{x}_U$,

then S or S_U and R or R_U don't change,

but the ith principal component is $\hat{\underline{e}}_i^T \underline{w}_j$

$= \hat{\underline{e}}_i (\underline{x}_j - T)$.

19) Classical PCA is affected by outliers, especially the 1st principal component.



data: x_j, $j = 1, \ldots, n$

Ch 7 1] p141 Let $\underline{x} = (\underline{w}^T\ \underline{y}^T)^T$ where

$\underline{w}$ is $m \times 1$, $m = p - q \le q$ and $\underline{Y}$ is $q \times 1$.

Canonical correlation analysis CCA seeks

m pairs of linear combinations $(\underline{a}_i^T \underline{w}, \underline{b}_i^T \underline{y})$

such that $\mathrm{corr}(\underline{a}_i^T \underline{w}, \underline{b}_i^T \underline{y})$ is large.

2] $\hat{U}_k = \hat{\underline{a}}_k^T \underline{w}$ and $\hat{V}_k = \hat{\underline{b}}_k^T \underline{y}$ are the

kth pair of canonical variables and the

kth canonical correlation $\hat{\rho}_k = \mathrm{corr}(\hat{U}_k, \hat{V}_k)$.

23.5

3) R output

cancor(X,Y) $\qquad$ ($\underline{x}$ was my $\underline{w}$)

$cor

$\hat{\rho}_1, \ldots, \hat{\rho}_m$

$X_1 \ldots X_m$ $\quad \underline{\hat{a}}_1, \ldots, \underline{\hat{a}}_m$ $\qquad Y_1 \ldots Y_q$ $\quad \underline{\hat{b}}_1, \ldots, \underline{\hat{b}}_m, \ldots, \underline{\hat{b}}_q$

4) Interpret $\underline{\hat{a}}_i$ and $\underline{\hat{b}}_i$ much as $\underline{\hat{e}}_j$ is interpreted for PCA.

ex) Let W_1 = arm length, $W_2, \ldots, W_m, Y_1, \ldots, Y_{q-1}$ iid $N(0,1)$ RVs independent of all other variables and Y_q = leg length. Want $\underline{a}^T \underline{w}$ and $\underline{b}^T \underline{Y}$ highly correlated with $\underbrace{\underline{a}^T \underline{a} = 1 \text{ and } \underline{b}^T \underline{b} = 1}_{\text{usually not true}}$.

Take $\underline{a} = (1, 0, \ldots, 0)^T$ and $\underline{b} = (0, \ldots, 0, 1)^T$.

5) p141 $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, $\quad \Sigma_a = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \quad (\lambda_i, \underline{a}_i)$

$\Sigma_A = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1/2} \quad (\lambda_i, \underline{e}_i)$

$\Sigma_b = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \quad (\lambda_i, \underline{b}_i)$

$\Sigma_B = \Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1/2} \quad (\lambda_i, \underline{g}_i)$.

For $i = 1, \ldots, m$ the m largest eigenvalues λ_i are the same, $\underline{e}_i$ and $\underline{g}_i$ are orthonormal,

$\underline{a}_i = \Sigma_{11}^{-1/2} \underline{e}_i$, $\quad \underline{b}_i = \Sigma_{22}^{-1/2} \underline{g}_i$.