PCA -- In practice

- The basic goal of PCA is to reduce the dimensionality of the data. Thus, one usually chooses:

$$n << q$$

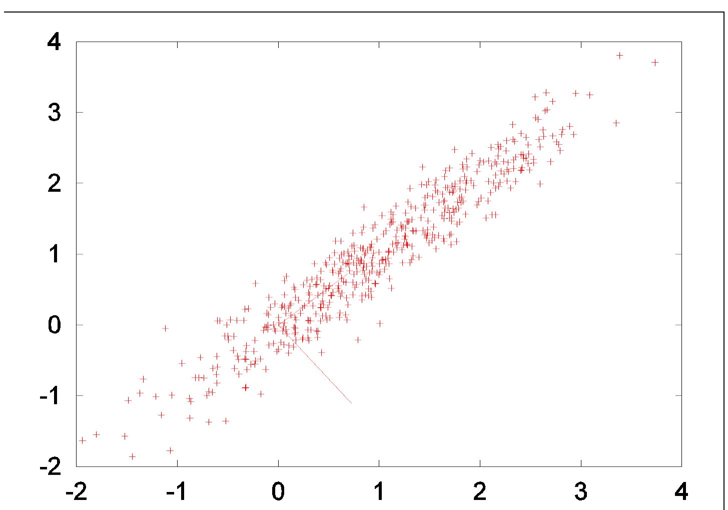
- But how do we select the number of components *n* ?

Determining the number of components

- Plot the eigenvalues – each eigenvalue is related to the amount of variation explained by the corresponding axis (eigenvector);
- If the points on the graph tend to level out (show an "elbow" shape), these eigenvalues are usually close enough to zero that they can be ignored.
- In general: Limit the variance accounted for.

Critical information lies in low dimensional subspaces

- A typical eigenvalue spectrum and its division into two orthogonal subspaces



$$\lambda_1 = 1.98, \quad \lambda_2 = 0.05$$

Determining the number of components

$$x_i \in \Re^q, \quad i = 1, \cdots, N$$

$w_1, w_2, \cdots, w_q$: q eigenvectors (principal component directions)

$\|w_i\| = 1$ (the w_is are orthonormal vectors)

Representation of x_i in eigenvector space:

$$y_i = \left(w_1^T x_i\right) w_1 + \left(w_2^T x_i\right) w_2 + \cdots + \left(w_q^T x_i\right) w_q$$

Suppose we retain the first k principal components:

$$y_i^k = \left(w_1^T x_i\right) w_1 + \left(w_2^T x_i\right) w_2 + \cdots + \left(w_k^T x_i\right) w_k$$

Then:

$$y_i - y_i^k = \left(w_{k+1}^T x_i\right) w_{k+1} + \cdots + \left(w_q^T x_i\right) w_q$$

Determining the number of components

$$\left(y_i - y_i^k\right)^T \left(y_i - y_i^k\right) =$$

$$\left[\left(w_{k+1}^T x_i\right) w_{k+1} + \cdots + \left(w_q^T x_i\right) w_q\right]^T \left[\left(w_{k+1}^T x_i\right) w_{k+1} + \cdots + \left(w_q^T x_i\right) w_q\right] =$$

$$w_{k+1}^T \left(w_{k+1}^T x_i\right)^2 w_{k+1} + \cdots + w_q^T \left(w_q^T x_i\right)^2 w_q =$$

(note $w_i^T w_j = 0 \ \forall i \neq j$ since w_i and w_j are orthogonal vectors)

$$\left(w_{k+1}^T x_i\right)^2 w_{k+1}^T w_{k+1} + \cdots + \left(w_q^T x_i\right)^2 w_q^T w_q =$$

$$\left(w_{k+1}^T x_i\right)^2 + \cdots + \left(w_q^T x_i\right)^2 =$$

$$\left(w_{k+1}^T x_i\right)\left(x_i^T w_{k+1}\right) + \cdots + \left(w_q^T x_i\right)\left(x_i^T w_q\right) =$$

$$w_{k+1}^T \left(x_i x_i^T\right) w_{k+1} + \cdots + w_q^T \left(x_i x_i^T\right) w_q$$

Determining the number of components

$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - y_i^k\right)^T\left(y_i - y_i^k\right) =$$

Mean square error

$$\frac{1}{N}\sum_{i=1}^{N}\left[\boldsymbol{w}_{k+1}^T\left(\boldsymbol{x}_i\boldsymbol{x}_i^T\right)\boldsymbol{w}_{k+1} + \cdots + \boldsymbol{w}_q^T\left(\boldsymbol{x}_i\boldsymbol{x}_i^T\right)\boldsymbol{w}_q\right] =$$

$$\boldsymbol{w}_{k+1}^T\left[\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{x}_i\boldsymbol{x}_i^T\right)\right]\boldsymbol{w}_{k+1} + \cdots + \boldsymbol{w}_q^T\left[\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{x}_i\boldsymbol{x}_i^T\right)\right]\boldsymbol{w}_q =$$

$$\boldsymbol{w}_{k+1}^T\Sigma\boldsymbol{w}_{k+1} + \cdots + \boldsymbol{w}_q^T\Sigma\boldsymbol{w}_q$$

We have: $\Sigma\boldsymbol{w}_{k+1} = \lambda_{k+1}\boldsymbol{w}_{k+1}, \cdots, \Sigma\boldsymbol{w}_q = \lambda_q\boldsymbol{w}_q$

Thus:

$$\boldsymbol{w}_{k+1}^T\Sigma\boldsymbol{w}_{k+1} + \cdots + \boldsymbol{w}_q^T\Sigma\boldsymbol{w}_q =$$

$$\boldsymbol{w}_{k+1}^T\lambda_{k+1}\boldsymbol{w}_{k+1} + \cdots + \boldsymbol{w}_q^T\lambda_q\boldsymbol{w}_q =$$

$$\lambda_{k+1} + \cdots + \lambda_q$$

Determining the number of components

$$\boxed{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - y_i^k\right)^T\left(y_i - y_i^k\right) = \lambda_{k+1} + \cdots \lambda_q}$$

The mean square error of the truncated representation is equal to the sum of the remaining eigenvalues.

In general: choose *k* so that 90-95% of the variance of the data is captured.