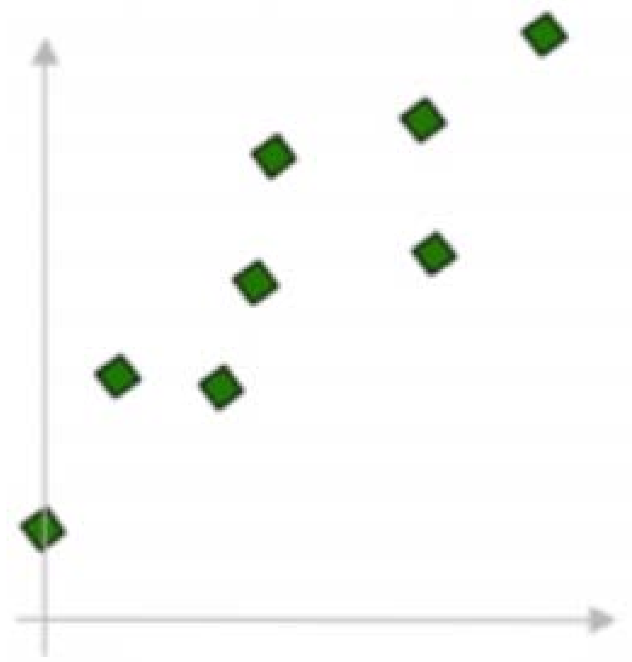
MACHINE LEARNING WITH PYTHON

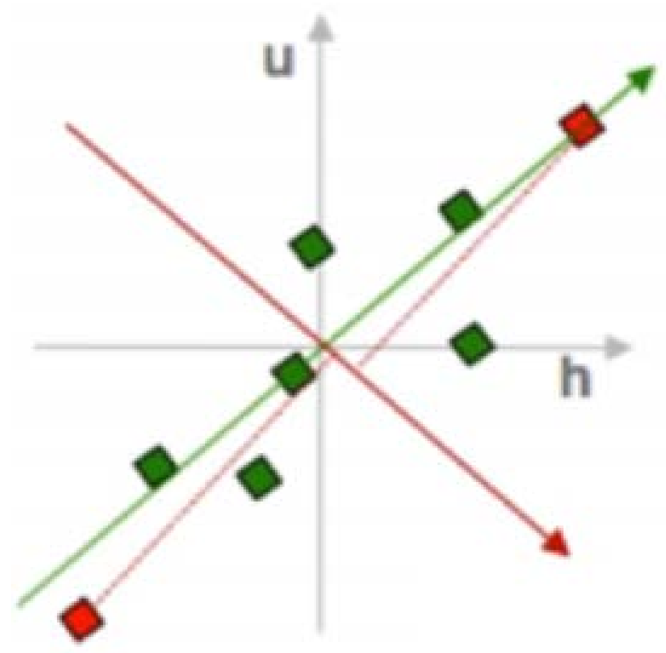
FEATURE EXTRACTION

Themistoklis Diamantopoulos

Principal Component Analysis

- Step 1: Center the data





Source: https://www.youtube.com/watch?v=QP43Iy-QQWY

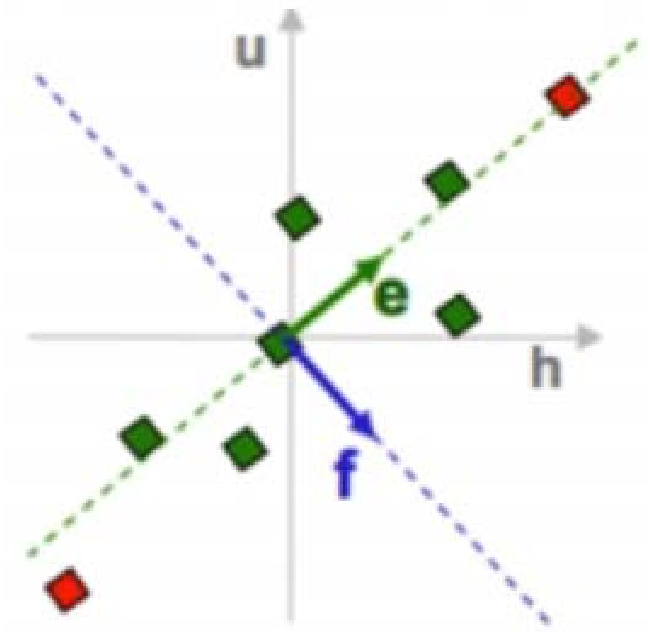
Principal Component Analysis

- Step 2: Find dimensions that maximize variance
  - Compute covariance matrix
  - Get eigenvalues and eigenvectors

$$\begin{array}{c c} & \begin{array}{cc} h & u \end{array} \\ \begin{array}{c} h \\ u \end{array} & \begin{pmatrix} 2.0 & 0.8 \\ 0.8 & 0.6 \end{pmatrix} \end{array} \rightarrow \operatorname{cov}(h,u) = \frac{1}{n}\sum_{i=1}^{n} h_i u_i$$

$$\begin{pmatrix} 2.0 & 0.8 \\ 0.8 & 0.6 \end{pmatrix} \begin{pmatrix} e_h \\ e_u \end{pmatrix} = \lambda_e \begin{pmatrix} e_h \\ e_u \end{pmatrix}$$

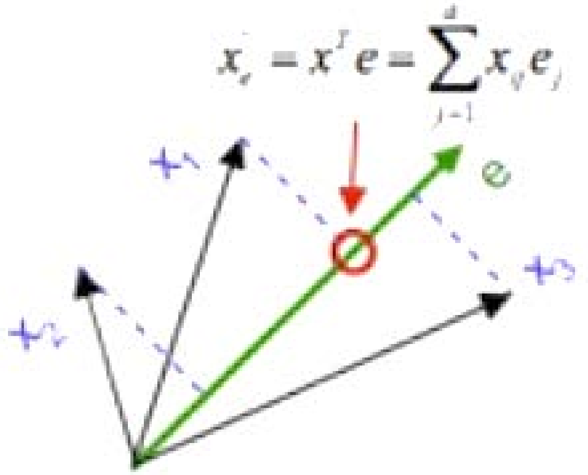
$$\begin{pmatrix} 2.0 & 0.8 \\ 0.8 & 0.6 \end{pmatrix} \begin{pmatrix} f_h \\ f_u \end{pmatrix} = \lambda_f \begin{pmatrix} f_h \\ f_u \end{pmatrix}$$

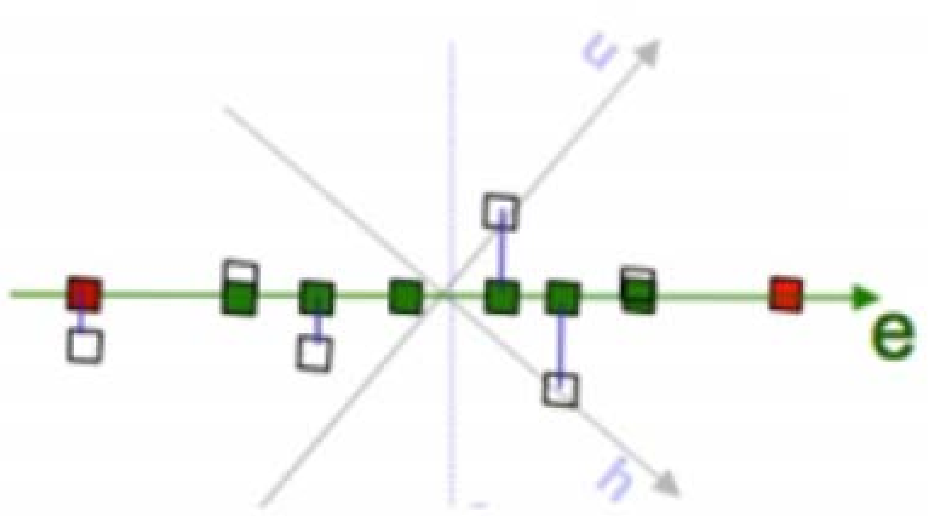


Source: https://www.youtube.com/watch?v=QP43Iy-QQWY

Principal Component Analysis

- Step 3: Project data points to PCs
  - Keep less dimensions





Source: https://www.youtube.com/watch?v=QP43Iy-QQWY