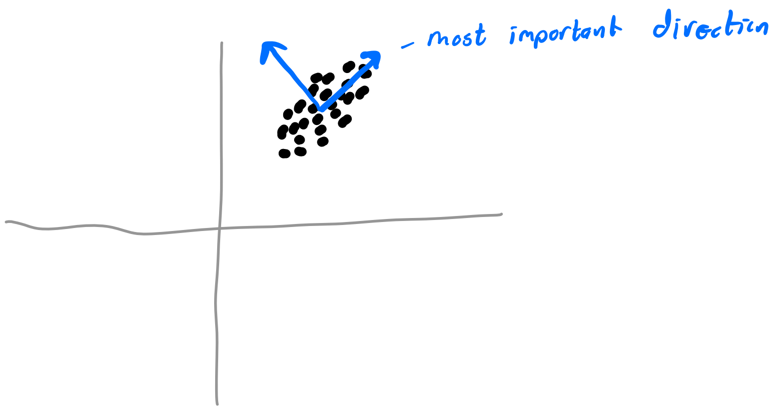
Day 20 - 17 November - Principal Component Analysis

Agenda:

- Principal Component Analysis
- Constrained Optimization
- Applications of PCA

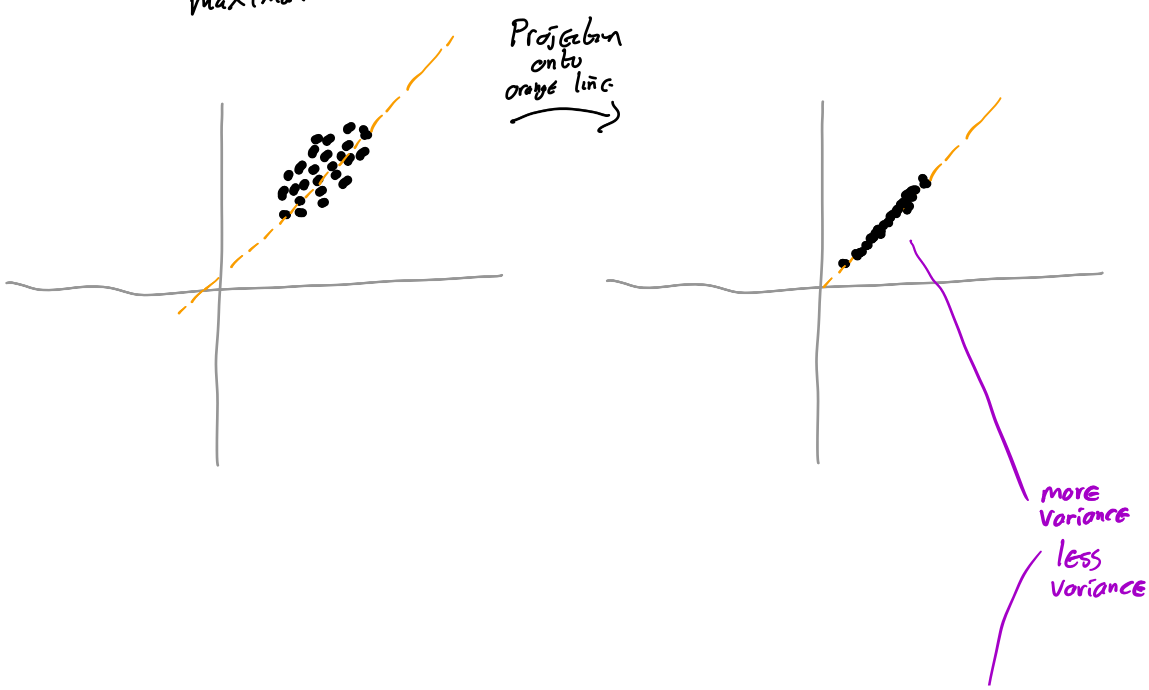
Principal Component Analysis (PCA)

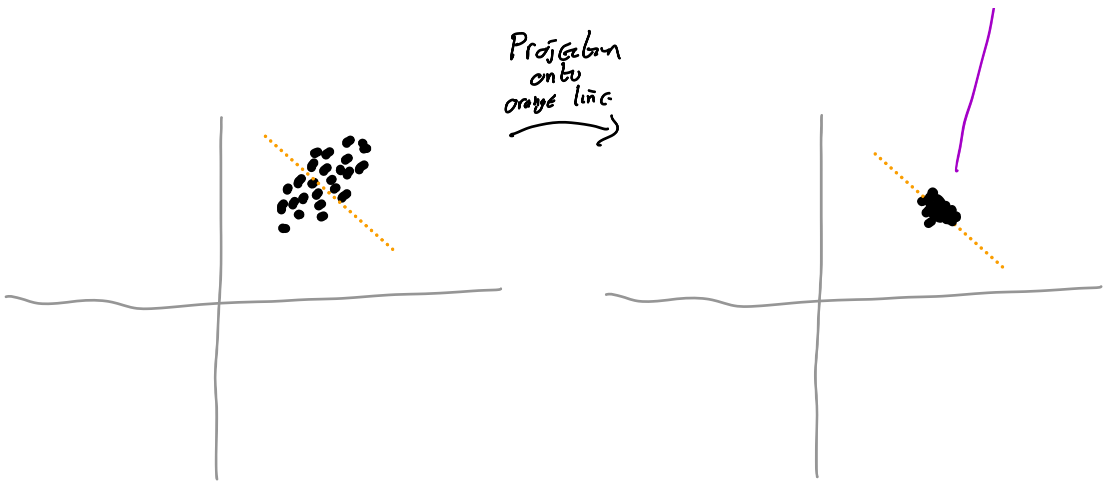
Given a set of data points $\{X_i\}_{i=1\cdots n}$ in $\mathbb{R}^d$, find the most important directions in $\mathbb{R}^d$ that explain the data



Maximum Variance Formulation of PCA

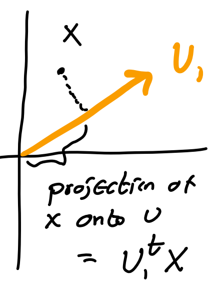
Goal: given $\{X_i\}_{i=1\cdots n}$, find subspace of dim M such that the variance of projected data is maximal





1d Case $(M=1)$

Find direction $u_1 \in \mathbb{R}^d$ such that variance of data projected on u_1 is largest. Note $u_1^t u_1 = 1$.



Each x_i projects to scalar value $u_1^t x_i$

Mean of $u_1^t x_i = u_1^t \bar{x}$ w/ $\bar{x} = \frac{1}{n}\sum_i x_i$

Variance of projected data

$$\frac{1}{n}\sum_{i=1}^{n}\left(u_1^t x_i - u_1^t \bar{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(u_1^t (x_i - \bar{x})\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} u_1^t (x_i - \bar{x})(x_i - \bar{x})^t u_1$$

$$= u_1^t \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^t\right)}_{S} u_1$$

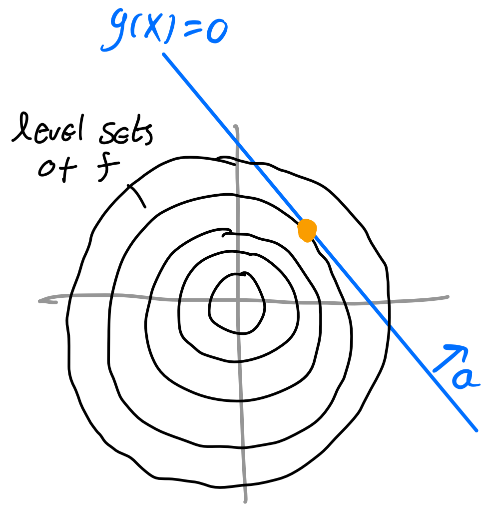
$$= u_1^t S u_1$$

data covariance matrix

Goal: $\underset{u_1}{\operatorname{argmax}} \; u_1^t S u_1 \quad \text{s.t.} \quad \|u_1\|^2 = 1$

Constrained Optimization

$$\min_{X} f(X) \quad st \quad g(X) = 0$$



Eg $\quad \min \|X\|^2 \quad st \quad \underset{\mathbb{R}^2}{a} \cdot \underset{\mathbb{R}^2}{X} = \underset{\mathbb{R}}{b}$

Find constrained optimizer by introducing Lagrange Multiplier

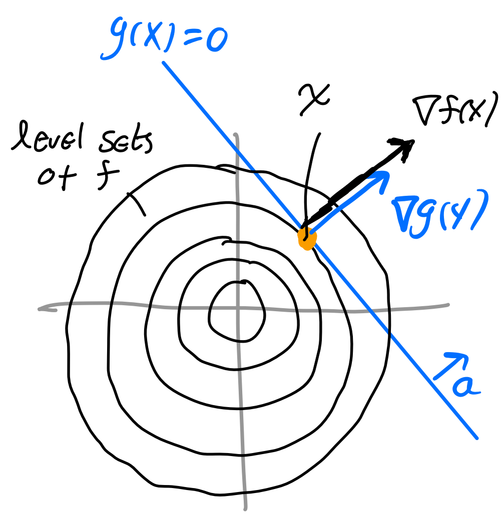
$$\min_{X} f(X) - \lambda g(X)$$

Now set gradient wrt X to 0

$$\nabla f(X) - \lambda \nabla g(X) = 0 \implies \nabla f(X) = \lambda g(Y)$$

Example:

$$\min \underbrace{X_1^2 + X_2^2}_{f(X)} \quad st \quad \underbrace{X_1 + X_2 = 1}_{g(X) = X_1 + X_2 - 1}$$



$$\min \quad X_1^2 + X_2^2 - \lambda (X_1 + X_2 - 1)$$

$$\nabla_X = \begin{pmatrix} 2X_1 - \lambda \\ 2X_2 - \lambda \end{pmatrix} = 0 \implies \begin{matrix} 2X_1 = \lambda \\ 2X_2 = \lambda \end{matrix}$$

$$\implies X_1 = X_2$$

So $\begin{matrix} X_1 + X_2 = 1 \\ \& \quad X_1 = X_2 \end{matrix} \implies \boxed{X_1 = X_2 = \frac{1}{2}}$

Back to 1d PCA

Goal:
$$\underset{U_1}{\operatorname{argmax}} \ U_1^t S U_1 \quad \text{s.t. } \|U_1\|^2 = 1$$

$$\max_{U_1} \ U_1^t S U_1 - \lambda (U_1^t U_1 - 1)$$

Taking $\nabla_{U_1} \cdot = 0$:

$$\nabla_{U_1} (U_1^t S U_1 - \lambda (U_1^t U_1 - 1))$$

$$= 2 S U_1 - \lambda U_1 = 0$$

$$\Rightarrow \quad S U_1 = \frac{\lambda}{2} U_1$$

$\Rightarrow$ U_1 is eigenvector of S.

So variance will be maximized at U_1 = eigenvector w/ largest eigenvalue.

Higher dim case ($M > 1$)

The M dimensional subspace on which the variance of the projected data is maximal is given by $\operatorname{span}(U_1 \cdots U_M)$ w/ U_i an eigenvector for i^{th} largest eigenvalue

The vectors $\{v_i\}$ are the principal components

Algorithm: Given $\{X_i\}$, M

Compute $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} \bar{X}$

$$S = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^t$$

Compute top M eigenvectors of S

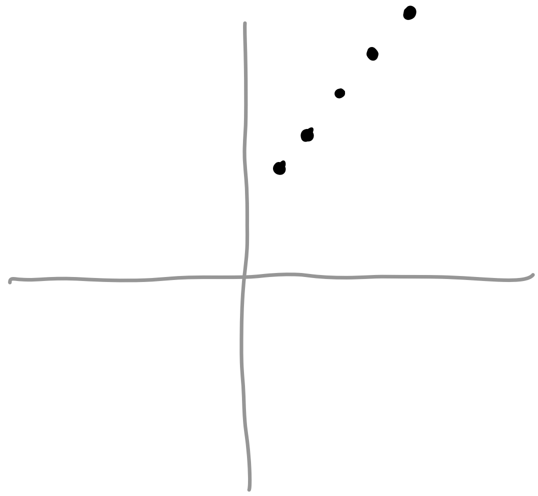
Output $\{v_1 \cdots v_M\}$

Cost: Computing a full eigenvalue decomposition of a $d \times d$ matrix takes $O(d^3)$ flops

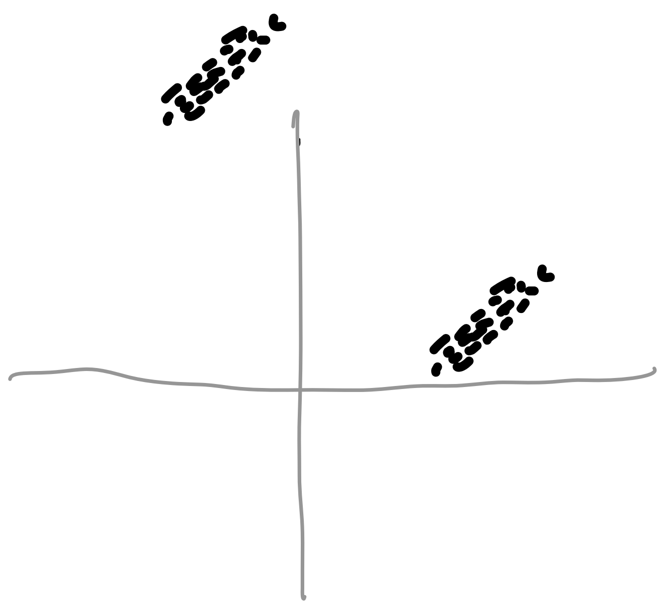
Computing just top M eigenvectors takes $O(Md^2)$ flops.

Illustrations:

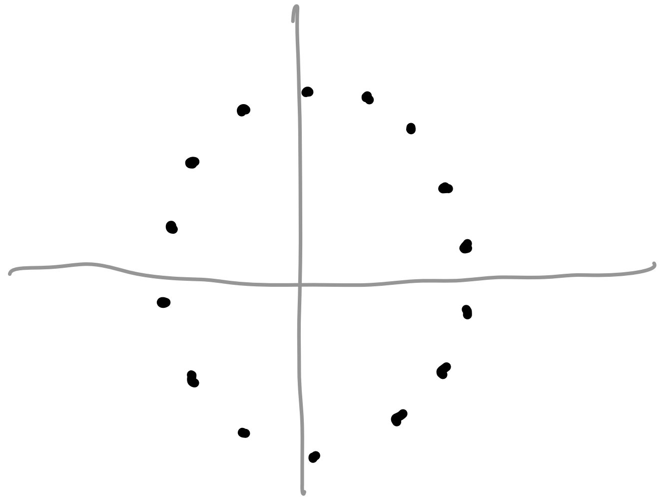
what will be the principal components of the following data?



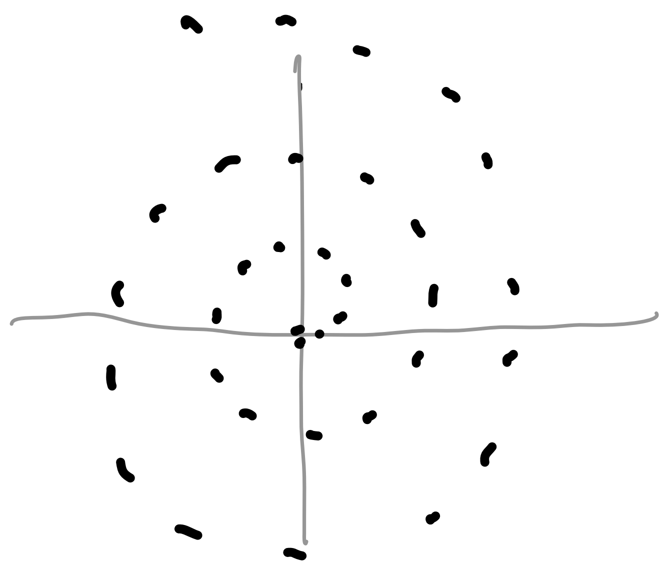
what will be the principal components of the following data?



What will be the principal components of the following data?



What will be the principal components of the following data?



Applications of PCA

dimensionality reduction
data preprocessing
compression (lossy)
data visualization

Whitening of data

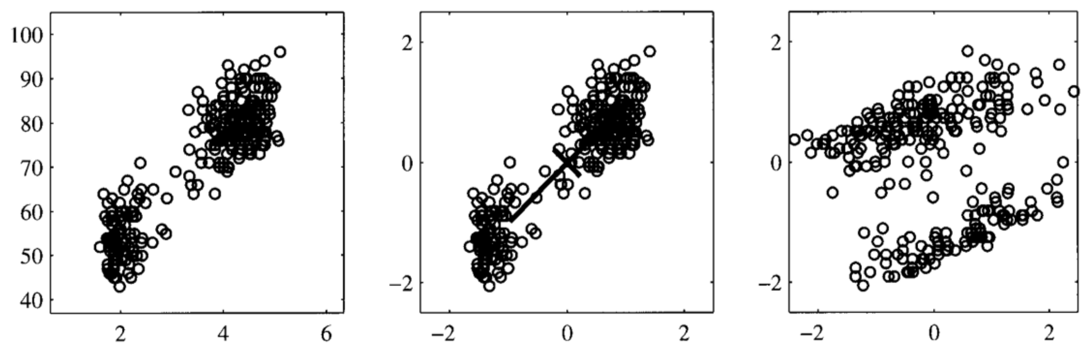


Figure 12.6 Illustration of the effects of linear pre-processing applied to the Old Faithful data set. The plot on the left shows the original data. The centre plot shows the result of standardizing the individual variables to zero mean and unit variance. Also shown are the principal axes of this normalized data set, plotted over the range $\pm\lambda_i^{1/2}$. The plot on the right shows the result of whitening of the data to give it zero mean and unit covariance.

Visualization

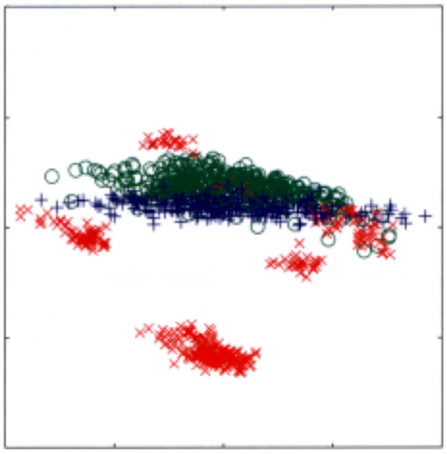


Figure 12.8 Visualization of the oil flow data set obtained by projecting the data onto the first two principal components. The red, blue, and green points correspond to the 'laminar', 'homogeneous', and 'annular' flow configurations respectively.