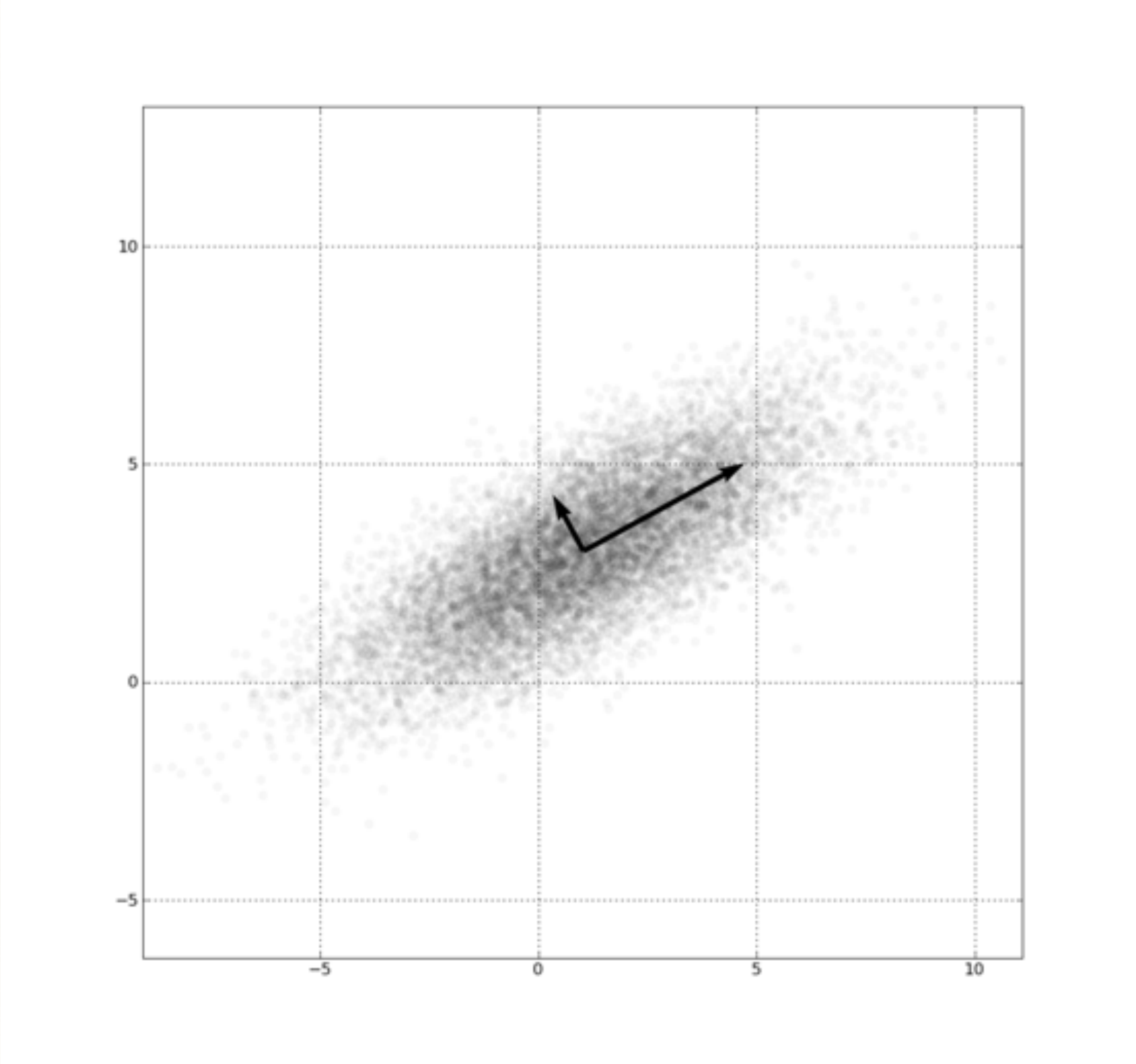
Michael Grossberg

Intro to Data Science CS59969

PCA / SVD Dimension Reduction

Dimension Reduction (Linear)

Principal Component Analysis (PCA)



1-D mean, stdev

Mean

$$\mu = E[x] = \frac{1}{n}\sum_{i=0}^{n} x_i$$

Variance = (Standard Deviation)^2

$$\sigma^2 = E[(x-\mu)^2] = \frac{1}{n}\sum_{i=0}^{n}(x_i-\mu)^2$$

Normal (1-D) distribution

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma}}$$

N-D mean, (co-)variance

N-D Mean

$$\mu = E[\mathbf{x}] = \frac{1}{n}\sum_{i=0}^{n} \mathbf{x}_i = \{\frac{1}{n}\sum_{k=0}^{n} x_{i,k}\}$$

N-D Covariance

$$\Sigma_{i,j} = \frac{1}{n}\sum_{k=0}^{n}(x_{i,k} - \mu)(x_{j,k} - \mu)$$

Matrix Versions

columns
dimensions / attributes

rows
observations

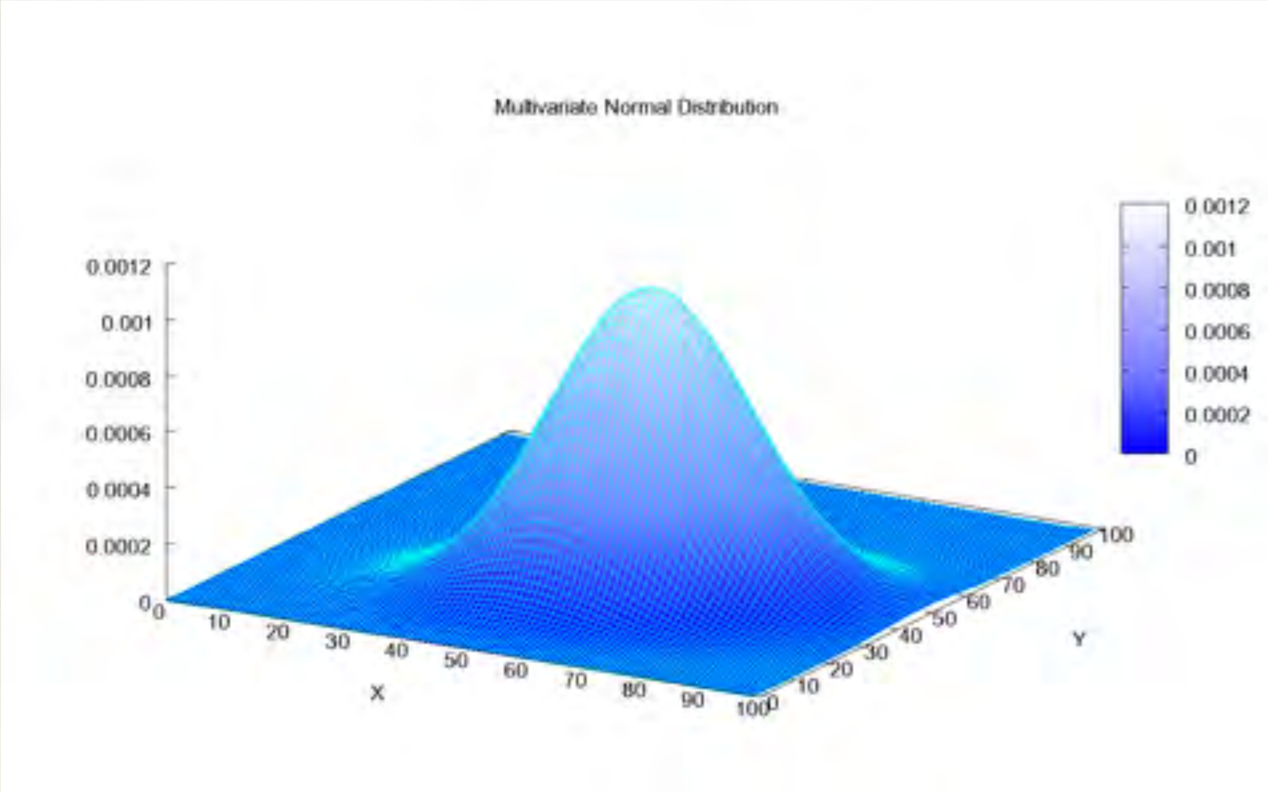
$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{n,1} \\ \vdots & \ddots & \vdots \\ x_{1,m} & \cdots & x_{n,m} \end{bmatrix}$$

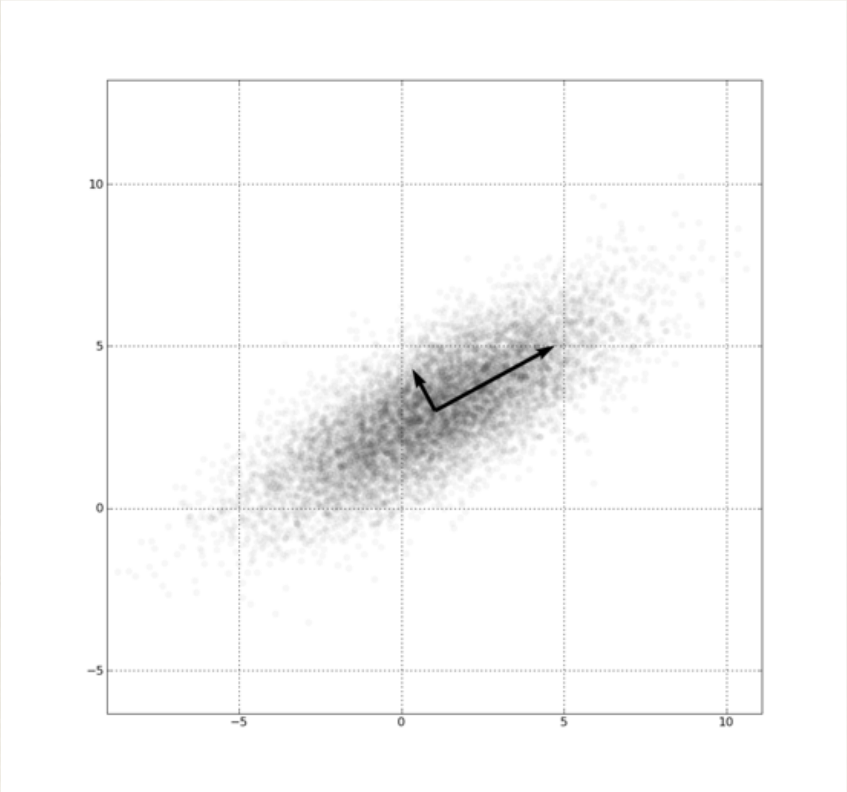
N-D Covariance

$$\Sigma = \check{X}^T \check{X}$$

N-D Normal Distribution

$$f_{\mathbf{x}}(x_1, \ldots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$





Diagonalization

$$\Sigma$$

Symmetric matrix

$$\Sigma = V^T D V$$

Diagonalization (SVD)

$$D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

Eigenvalues
Variances

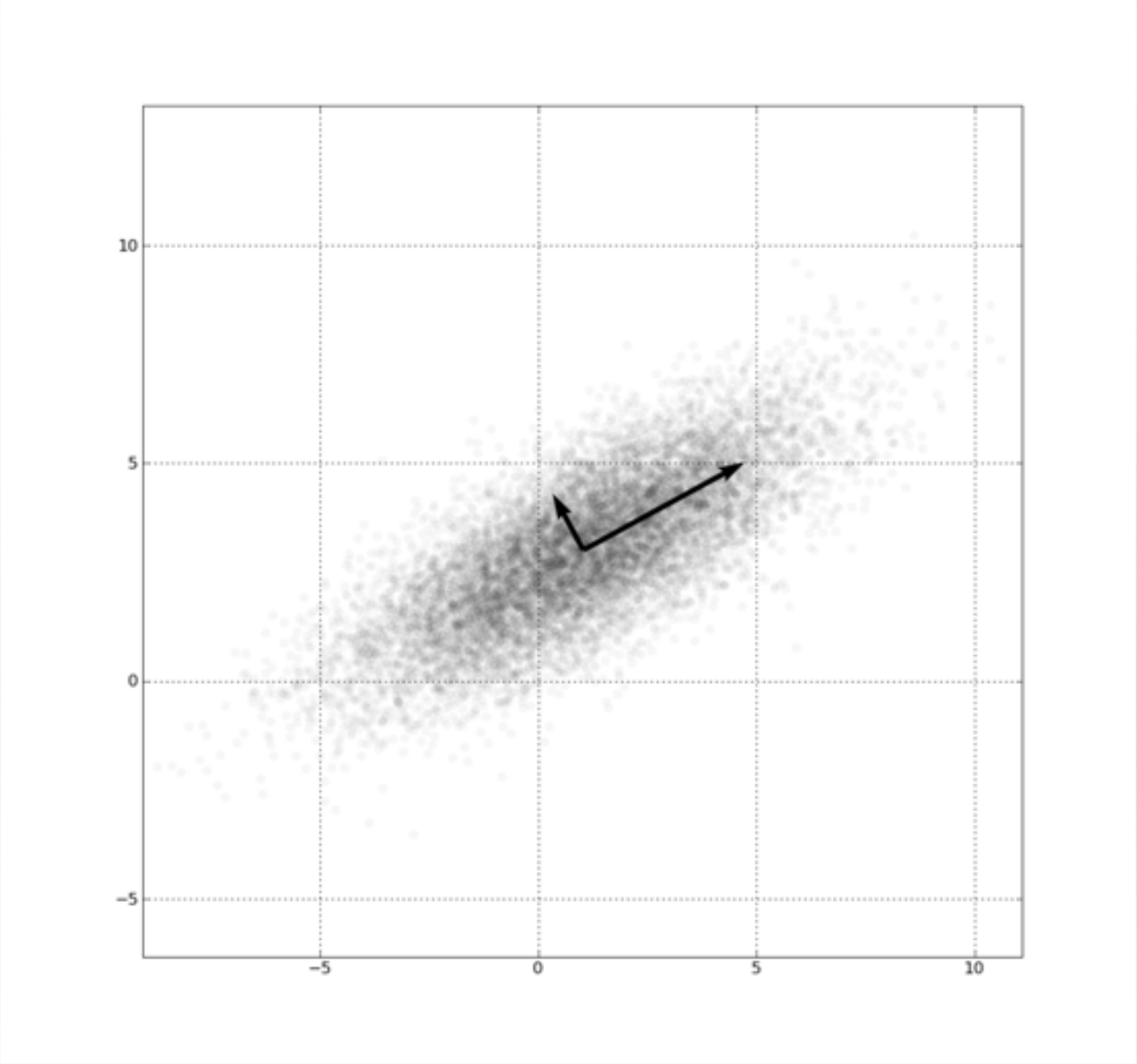
$$\lambda_1 > \lambda_2 \cdots > \lambda_n$$

$$V = \begin{bmatrix} \boxed{v_{1,1} \quad \cdots \quad v_{n,1}} \\ \vdots \quad \ddots \quad \vdots \\ \boxed{v_{1,n} \quad \cdots \quad v_{n,n}} \end{bmatrix}$$

First principal component

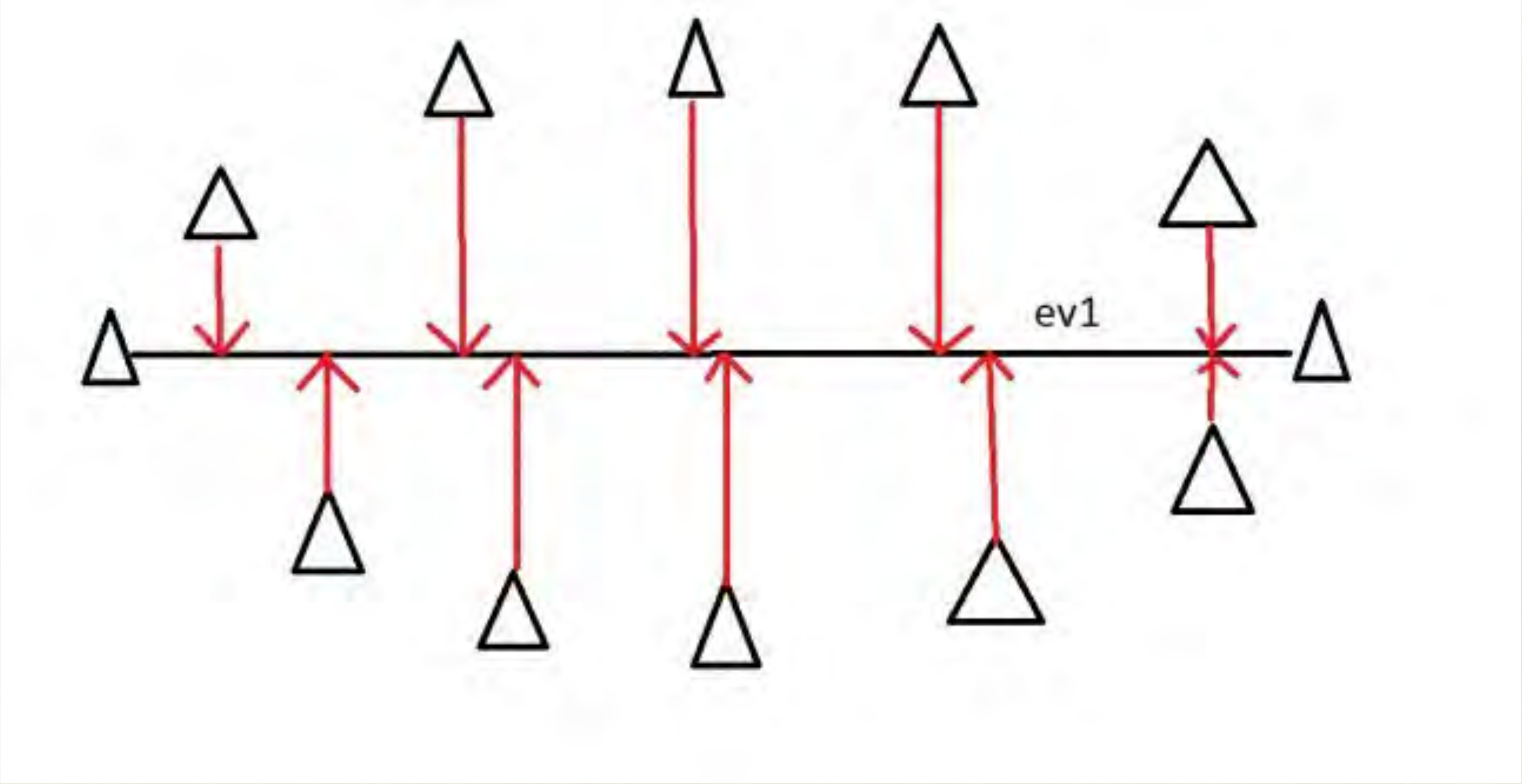
Eigenvectors

Eigenvectors = Principal Components

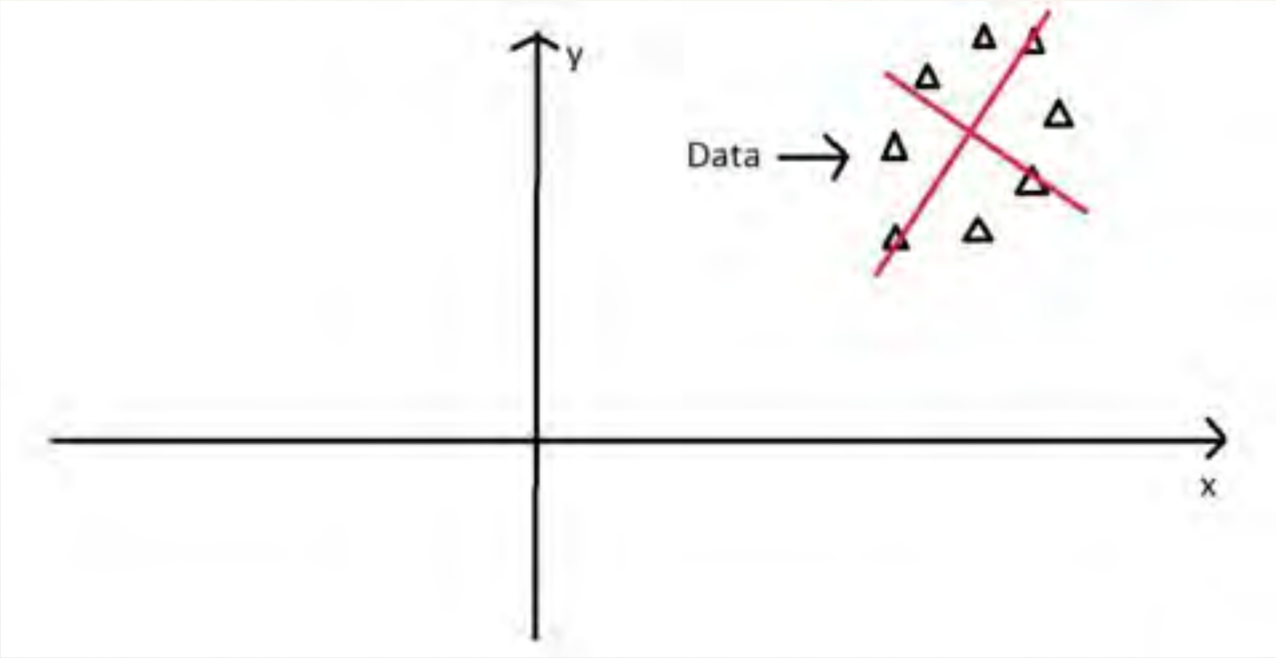


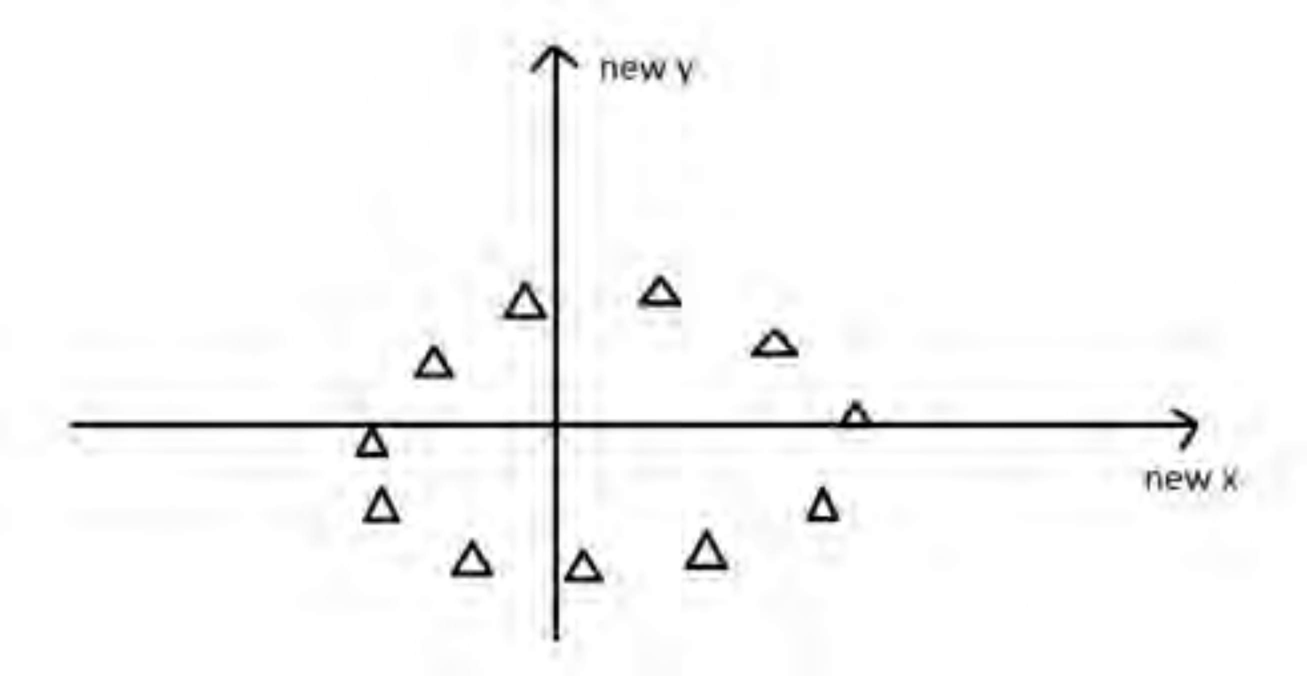
First principal component = direction of greatest variance

Least Squares



Data Oriented Coordinates





http://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/