Warm-up

1. https://www.sporcle.com/games/MrChewypoo/minimalist_disney

2. https://www.sporcle.com/games/Stanford0008/minimalist-cartoons-slideshow

3. https://www.sporcle.com/games/MrChewypoo/minimalist

Announcements

HW3

- Due Sat 10/9

Quiz3

- Mon 10/11, last 15 min. of class (probability, statistics)

Next week

- Mon 10/11: Review session
- Wed 10/13: No class
- Thu 10/14: Final exam, 1-4 pm

Mathematical Foundations for Machine Learning

Probability/Statistics & PCA Application

Instructor: Pat Virtue

Plan

Today

Probability/Statistics

- Expectation
- Variance
- Covariance
- Covariance matrix

10-606 Application: PCA

Expectation and Variance

Expectation and Variance

The **expected value** of X is $E[X]$. Also called the mean.

- Discrete random variables:

  Suppose X can take any value in the set $\mathcal{X}$.

  $$E[X] = \sum_{x \in \mathcal{X}} x p(x)$$

- Continuous random variables:

  $$E[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

Slide credit: CMU MLD Matt Gormley

Expectation and Variance

The **variance** of *X* is *Var(X)*.

$$Var(X) = E[(X - E[X])^2]$$

$\mu = E[X]$

- Discrete random variables:

$$Var(X) = \sum_{x \in \mathcal{X}} (x - \mu)^2 p(x)$$

- Continuous random variables:

$$Var(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$$

Covariance

Expectation: $E[x] = \sum_x x\, p(x)$

Variance: $Var[x] = E[(X - E(x))^2]$

Covariance: $Cov[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])]$

Vector of random variables: $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$

Covariance matrix:

$$\Sigma = \begin{bmatrix} Var[X_1] & Cov[X_1, X_2] & \cdots \\ Cov[X_2, X_1] & \ddots & \\ \vdots & & \ddots \\ Cov[X_N, X_1] & & \end{bmatrix}$$

Principal Component Analysis

Warm-up as you log in

1. https://www.sporcle.com/games/MrChewypoo/minimalist_disney

2. https://www.sporcle.com/games/Stanford0008/minimalist-cartoons-slideshow

3. https://www.sporcle.com/games/MrChewypoo/minimalist

Dimensionality Reduction

Dimensionality Reduction

Dimensionality Reduction

For each $x^{(i)} \in \mathbb{R}^M$ find representation $z^{(i)} \in \mathbb{R}^K$ where $K \ll M$

Dimensionality Reduction

http://timbaumann.info/svd-image-compression-demo/

https://cs.stanford.edu/people/karpathy/convnetjs/demo/autoencoder.html

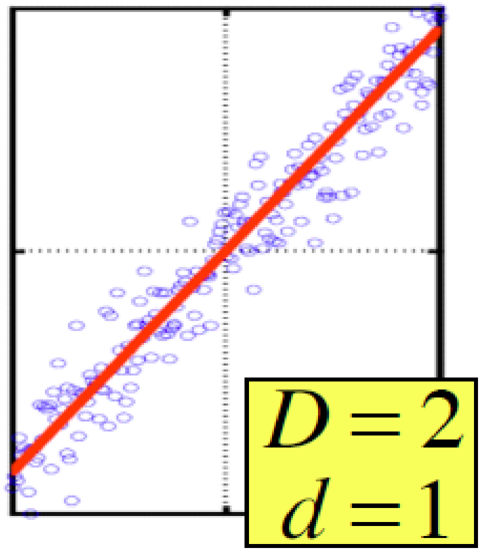
PCA Worksheet

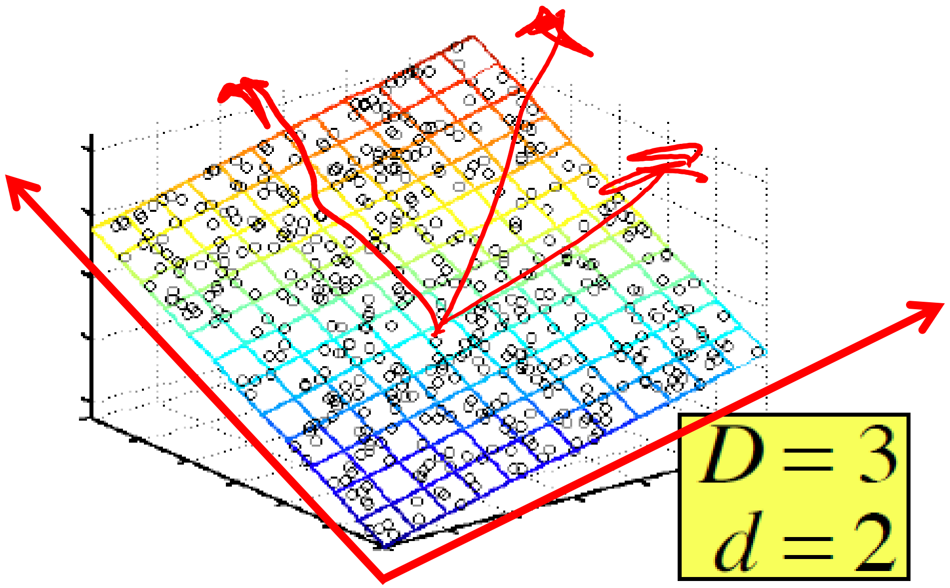
Lagrange multipliers for PCA

$$\hat{\mathbf{v}} = \underset{\mathbf{v}}{\text{argmax.}} \ \mathbf{v}^T A \mathbf{v}$$

$$\text{s.t.} \ \|\mathbf{v}\|_2^2 = 1$$

Principal Component Analysis (PCA)

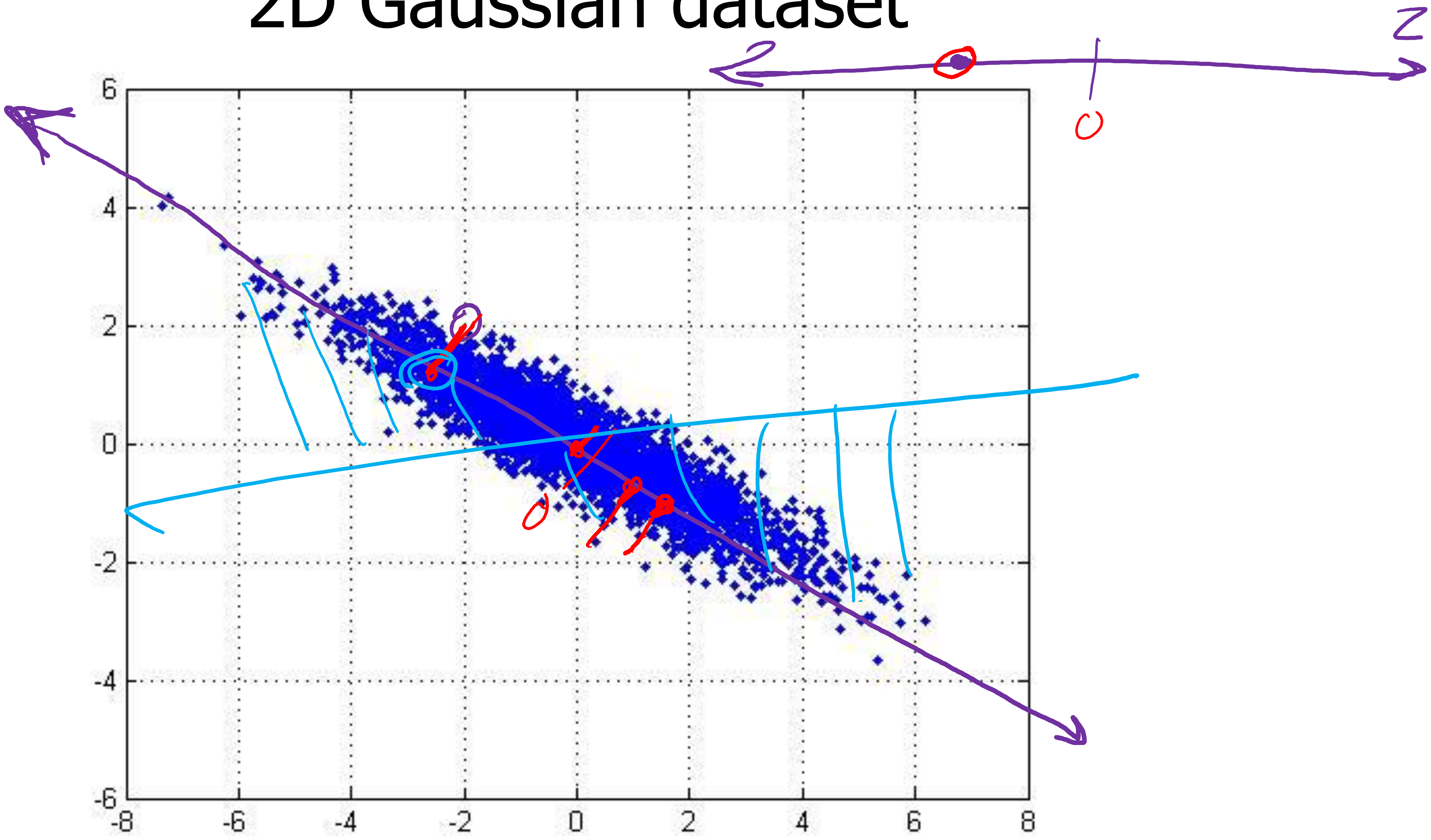




In case where data lies on or near a low d-dimensional linear subspace, axes of this subspace are an effective representation of the data.

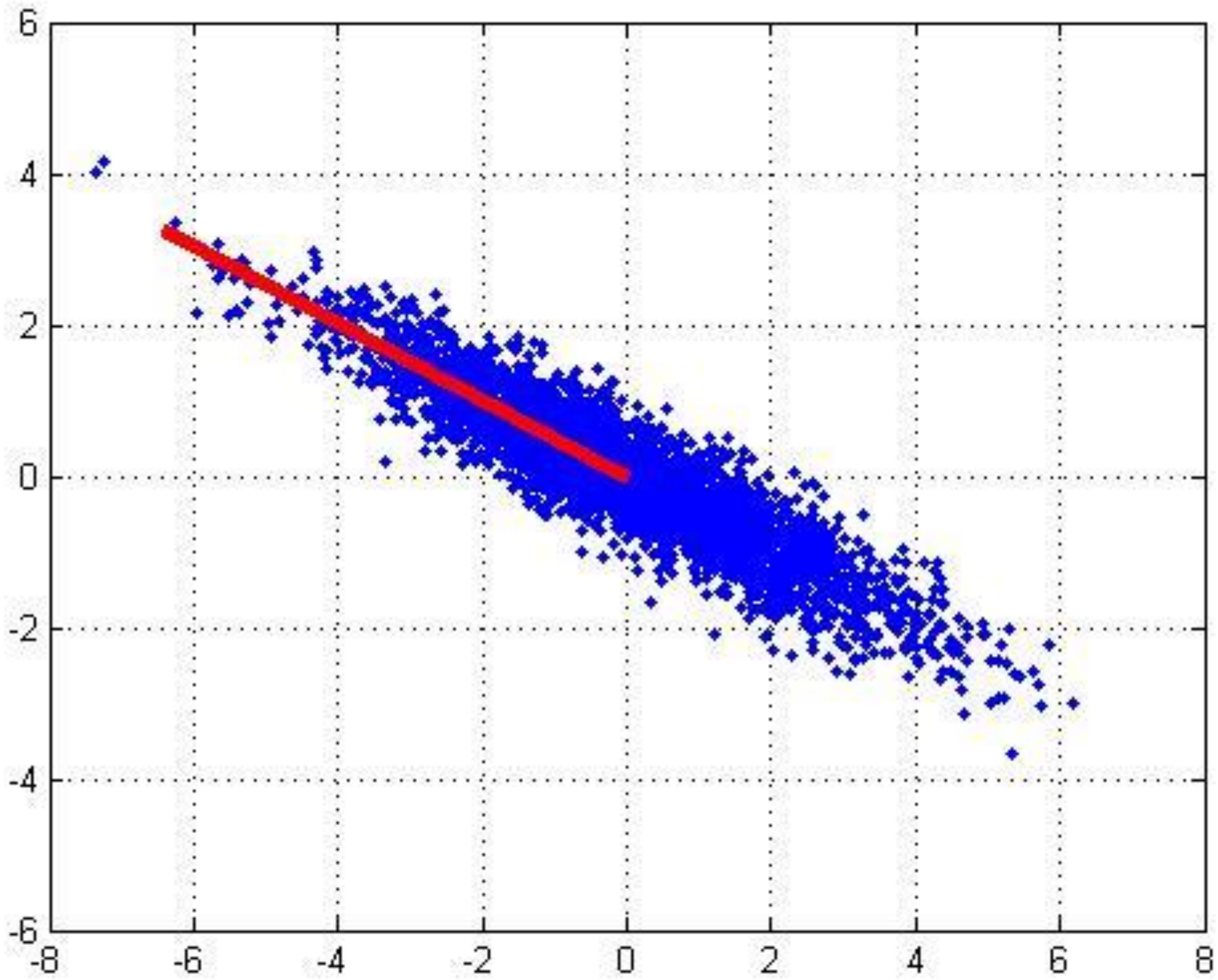
Identifying the axes is known as Principal Components Analysis, and can be obtained by using classic matrix computation tools (Eigen or Singular Value Decomposition).

2D Gaussian dataset



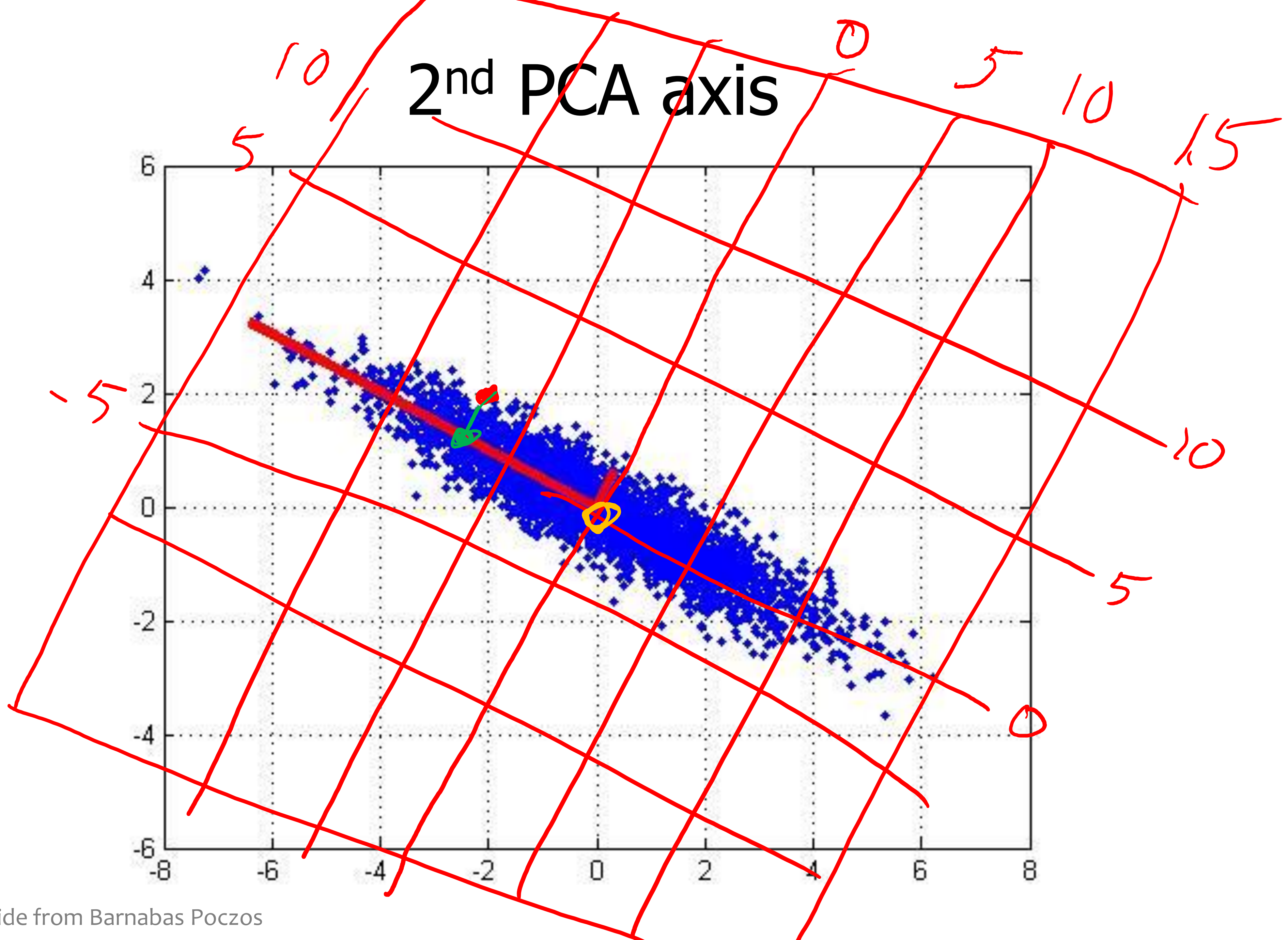
Slide from Barnabas Poczos

1st PCA axis



Slide from Barnabas Poczos

2nd PCA axis



Slide from Barnabas Poczos

Data for PCA

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{N}$$

$$\mathbf{X} = \begin{bmatrix} (\mathbf{x}^{(1)})^T \\ (\mathbf{x}^{(2)})^T \\ \vdots \\ (\mathbf{x}^{(N)})^T \end{bmatrix}$$

We assume the data is **centered**

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}^{(i)} = \mathbf{0}$$

Q: What if your data is **not** centered?

A: Subtract off the sample mean

Slide from Matt Gormley

Sample Covariance Matrix

The sample covariance matrix is given by:

$$\Sigma_{jk} = \frac{1}{N} \sum_{i=1}^{N} (x_j^{(i)} - \mu_j)(x_k^{(i)} - \mu_k)$$

Since the data matrix is centered, we rewrite as:

$$\mathbf{\Sigma} = \frac{1}{N} \mathbf{X}^T \mathbf{X}$$

$$\mathbf{X} = \begin{bmatrix} (\mathbf{x}^{(1)})^T \\ (\mathbf{x}^{(2)})^T \\ \vdots \\ (\mathbf{x}^{(N)})^T \end{bmatrix}$$

Slide from Matt Gormley

PCA Algorithm

Input: $\boldsymbol{X}, \boldsymbol{X}_{test}, K$

1. Center data (and scale each axis) based on training data → $\boldsymbol{X}, \boldsymbol{X}_{\boldsymbol{test}}$
2. $\boldsymbol{V}$ = eigenvectors($\boldsymbol{X}^T\boldsymbol{X}$)
3. Keep only the top K eigenvectors: $\boldsymbol{V}_K$
4. $\mathbf{Z}_{\text{test}} = \boldsymbol{X}_{test}\boldsymbol{V}_K$

Optionally, use $\boldsymbol{V}_K^T$ to rotate $\mathbf{Z}_{\text{test}}$ back to original subspace $\mathbf{X'}_{\text{test}}$ and uncenter