Kernel Q+A

Goals: Q+A, HW
Kernel PCA

Next week: Quiz early in week

Project released soon ← 2 datasets
allowed to use any software

Grad proposal: due next Fri

HW A5 due Wed

Office hrs. Mon instead of Fri

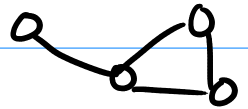
$$\vec{x} \longrightarrow \phi(\vec{x})$$

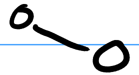
$$\vec{x}^T \vec{x}\,' \longrightarrow \underbrace{\vec{\phi}(\vec{x})^T \vec{\phi}(\vec{x}\,')}_{K(\vec{x},\,\vec{x}\,')}$$

two different data pts $\vec{x} \in \mathbb{R}^d$

ex/ $K(\vec{x},\vec{x}\,') = \exp\left(-\frac{\|\vec{x}-\vec{x}\,'\|^2}{2\sigma^2}\right)$

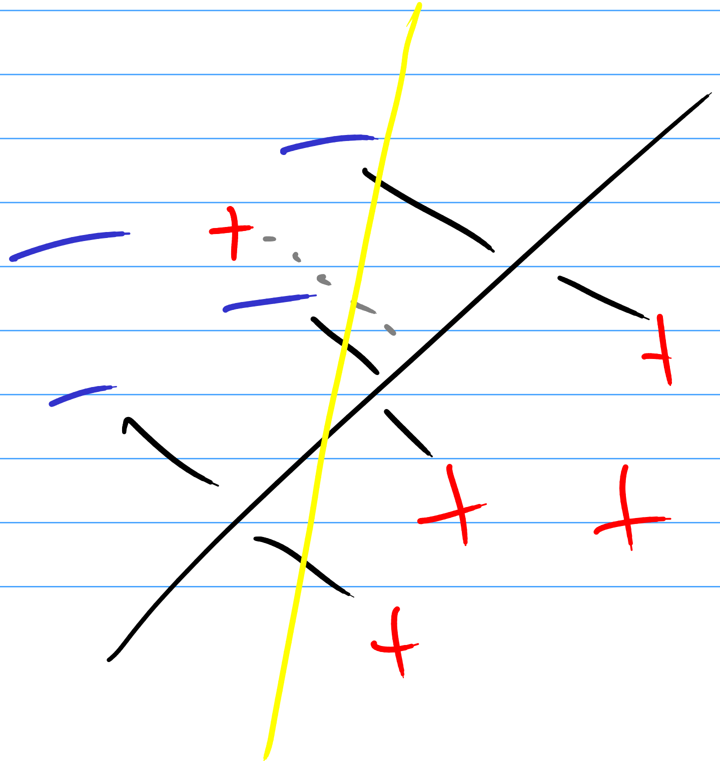
RBF kernel

$\vec{x}$ could be a graph  or string "ftp://hi"

$\vec{x}\,'$  "http://google"

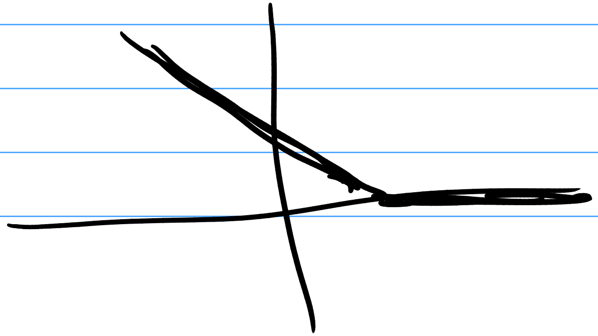
KRR : $\min_{\vec{a}} \|K\vec{a}-\vec{y}\|^2 + \lambda \vec{a}^T K \vec{a} \longrightarrow \vec{a} = (K+\lambda I)^{-1}\vec{y}$

SVM : $\min_{f} \sum_{i=1}^{n} \ell_{hinge}\left(y_i f(\vec{x}_i)\right) + \frac{1}{C}\underbrace{\|f\|_{\mathcal{H}}^2}_{\vec{a}^T K \vec{a}}$



$\min \ \vec{a}^T K \vec{a}$

s.t. $y_i f(\vec{x}_i) \geq m - s_i$



$$\vec{x} \longrightarrow \underbrace{\phi(\vec{x})}_{\text{function of } (\cdot)} = K(\vec{x}, \underset{\substack{\uparrow \\ \text{free variable}}}{\cdot})$$

reproducing property

PCA We computed eigenvalues/eigenvectors of a centered data matrix

$$\bar{X} = X - \text{column means} = \underbrace{\left(I - \frac{\vec{1}\vec{1}^T}{n}\right)}_{\text{matrix w/ all entries} = 1/n} X$$

(used SVD)

$$\bar{K} = \bar{X}\bar{X}^T = U D^2 U^T$$

" centered kernel " — $n \times n$ matrix of covariances — eigendecomposition

$$\bar{K} = \left(I - \frac{\vec{1}\vec{1}^T}{n}\right) \underbrace{X X^T}_{n \times n \text{ matrix of dot products}} \left(I - \frac{\vec{1}\vec{1}^T}{n}\right) = \left(I - \frac{\vec{1}\vec{1}^T}{n}\right) K \left(I - \frac{\vec{1}\vec{1}^T}{n}\right)$$

$$\frac{1}{n}\vec{1}\vec{1}^T \overset{(n=4)}{=} \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\frac{1}{n}\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \underbrace{[1 \dots 1]}_{n \text{ many}}$$

$\bar{K}$ is matrix of kernel features that have been centered

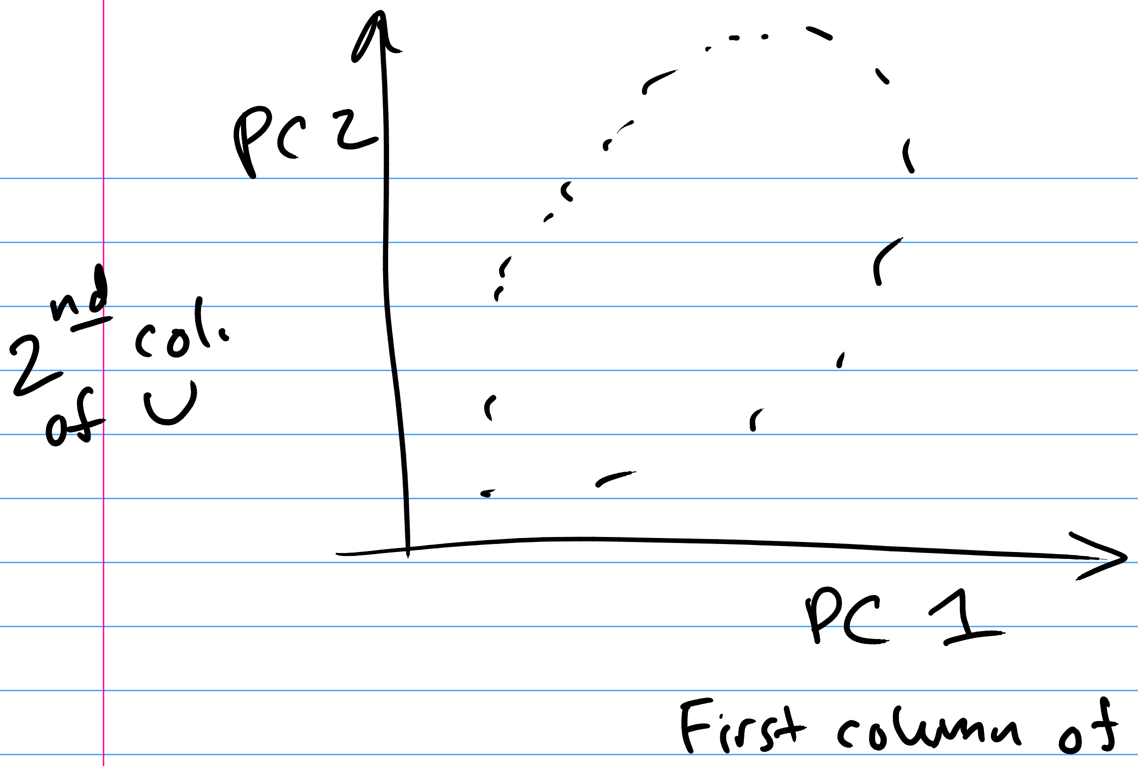
Alg:

Get data $\{\vec{x}_i\}_{i=1}^n$

Form K $n\times n$ matrix $K_{ij} = k(\vec{x}_i, \vec{x}_j)$

Form $\bar{K}$ using formula

Compute $\bar{K} = UD^2U^T$ eigendecomp.

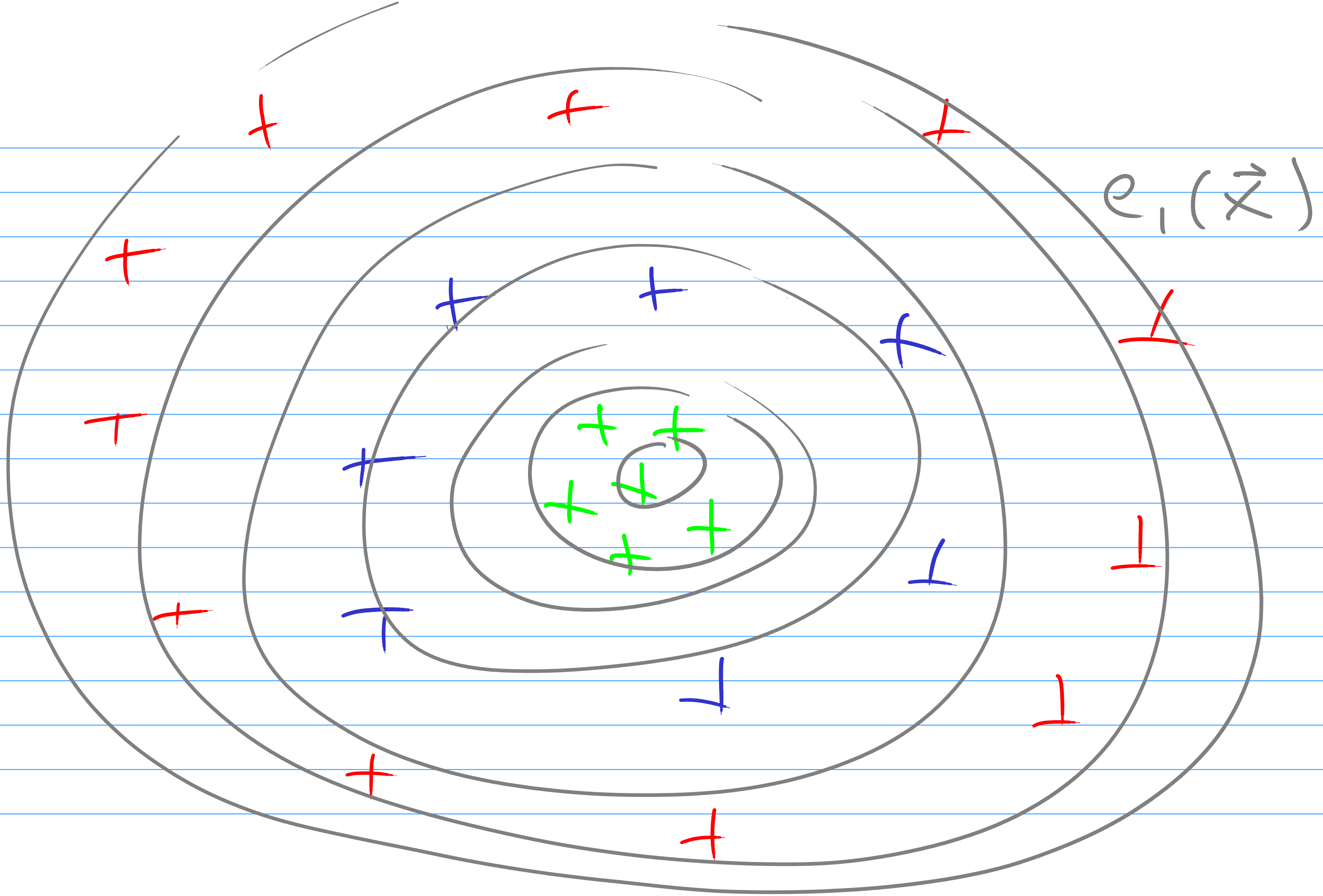


PCA : projection was linear

KPCA : projection is nonlinear

From U, D can compute eigenfunction $e_i(\vec{x})$ for each principal component

$$e_i(\vec{x}_j) = U_{ji}$$



$e_1(\vec{x})$

Poll

3 More unsupervised
nonneg. matrix tensor factorization ←

6 Multiclass classifications

8 More neural networks ←

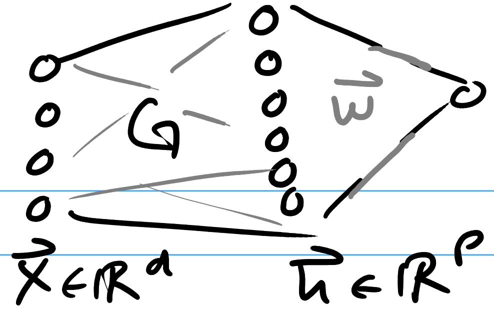
12 theory of learning/generalization

#3

applied entry-wise to p vector

$$\vec{h} = \cos\left(G\vec{x} + \vec{c}\right)$$

length p — p×d matrix — p vector



picking σ
variance of entries in G

Splitting

$X_{train} \longrightarrow$ training + validation
80% 20%

G for each σ

evaluate error

3.3 first time you touch