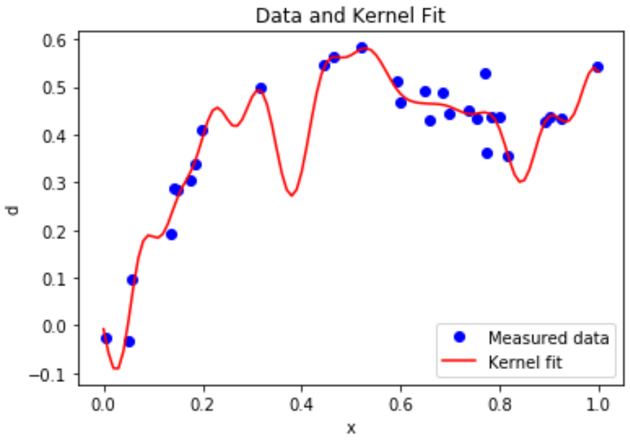
CS/ECE/ME 532
Period 23

Today

- Kernel methods: video 6.3

Applications:

- *Regression*



- *Classification (Kernel SVMs)*

- *Nothing to do with kernel of a matrix:*

$$\ker(\boldsymbol{A}) \text{ is all vectors such that } \boldsymbol{A}\boldsymbol{x} = 0$$

Wikipedia:

Kernel may refer to:

Computing [edit]

- Kernel (image processing), a matrix used for image conv
- Kernel (operating system), the central component of mos
- Compute kernel, in GPGPU programming
- Kernel method, in machine learning
- In numerical analysis, a subroutine that performs a comm
  - In particular, a routine that is executed in a vectorizec
- Kernelization, a technique for designing efficient algorithr

Mathematics [edit]

Objects [edit]

- Kernel (algebra), a general concept that includes:
  - Kernel (linear algebra), the set of all vectors which ma
  - Kernel (category theory), in category theory
  - Kernel (set theory), the set of all pairs of elements tha
  - Equalizer (mathematics), the set of all elements wher
- Kernel of a directed graph, a subset of the vertex set of a

Functions [edit]

- Kernel (geometry), the set of points within a polygon from
- Kernel (statistics), a weighting function used in kernel de
- Integral kernel, a function of two variables that defines ar
- Heat kernel, the fundamental solution to the heat equatio
- Convolution kernel
- Stochastic kernel, the transition function of a stochastic p
- Transition kernel, a generalization of a stochastic kernel
- Pricing kernel, the stochastic discount factor used in mat
- Positive-definite kernel, a generalization of a positive-def
- Kernel trick, in statistics
- Reproducing kernel Hilbert space

Kernels (in Machine Learning)

Binary classification: $\widehat{y} = \text{sign}(\boldsymbol{x}^T \boldsymbol{w})$

Linear regression: $\widehat{y} = \boldsymbol{x}^T \boldsymbol{w}$

Linear regression, after feature map:

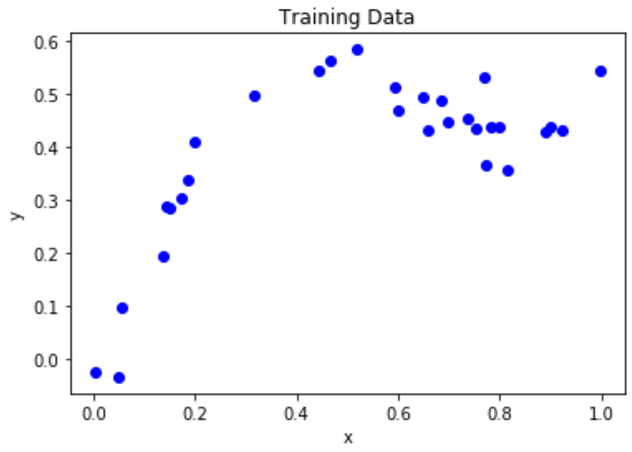
$$\widehat{y} = \phi(\boldsymbol{x})^T \boldsymbol{w}$$

$\boldsymbol{w}$ depends on $\boldsymbol{x}_1, y_1, \boldsymbol{x}_2, \boldsymbol{y}_2 \ldots$

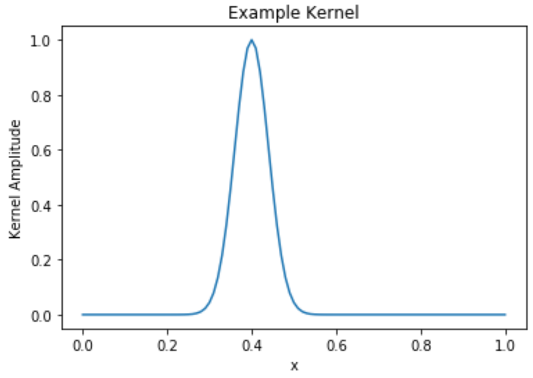
Kernel methods – re-write above as:

$$\widehat{y} = \sum_i \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i)$$

weighted sum of similarities between feature vector and each training point



$K(\boldsymbol{x}, \boldsymbol{x}_i) = e^{-||\boldsymbol{x}-\boldsymbol{x}_i||^2}$



$\boldsymbol{x} = [0.4]$ How do we predict $\hat{y}$? $\widehat{y} = \sum_i \alpha_i e^{-|0.4-x_i|^2}$

How do we find good α_i?

start by finding $\boldsymbol{w}$ using ridge regression

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} ||\boldsymbol{\Phi}\boldsymbol{w} - \boldsymbol{y}|| + \lambda||\boldsymbol{w}||^2$$

$$\boldsymbol{w}^* = (\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \lambda \boldsymbol{I})^{-1}\boldsymbol{\Phi}^T\boldsymbol{y}$$

$$\widehat{y} = \phi(\boldsymbol{x})^T \boldsymbol{w}^* \qquad \widehat{y} = \sum_i \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i)$$

Manipulations/comparisons

$$\boldsymbol{\alpha} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^T + \lambda \boldsymbol{I})^{-1}\boldsymbol{y}$$

where $\boldsymbol{\Phi}\boldsymbol{\Phi}^T$ has i, j entry $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$

No need to compute $\phi(\cdot)$ to compute $K(\cdot, \cdot)$ or $\widehat{y}$!