CS/ECE/ME 532
Matrix Methods in Machine Learning

Welcome!

Announcements

- Assignments:
  - Assignment 9 is due on Wednesday, May 1
  - Assignment 10 is due on *Friday*, May 3
  - No extensions!
  - **We will drop the *two* lowest Assignments (instead of one)**

- **Assessment 4 will be on Tuesday, May 7, at 5:35 pm**
  - If you have a conflict, you must let me know by Thursday, April 25 (fill out “Request for taking Assessment 4 at a different time”)

Activity 21

Recall

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

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

Kernels (in Machine Learning)

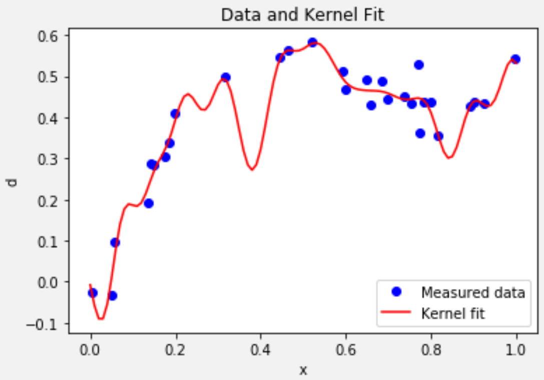
$\boldsymbol{w}$ depends on $\boldsymbol{x}_1, y_1, \boldsymbol{x}_2, \boldsymbol{y}_2 \dots$

Kernel

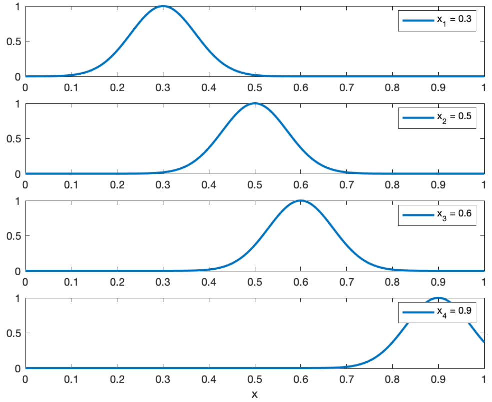
$$\widehat{y} = \boldsymbol{\phi}(\boldsymbol{x})^T \boldsymbol{w} \quad \longrightarrow \quad \widehat{y} = \sum_i \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i)$$

High dimensional

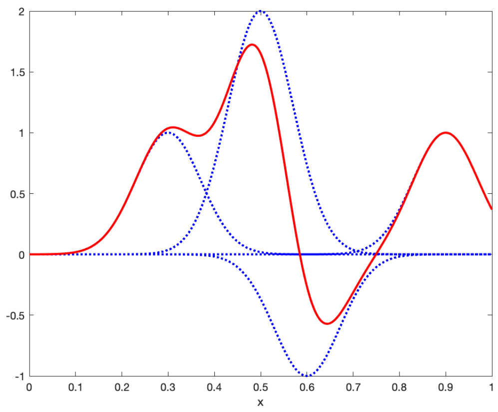
weighted sum of similarities between feature vector and each training point



$$K(x, x_i) = e^{-\frac{(x - x_i)^2}{0.01}}$$



$$\hat{y}(x) = K(x, x_1) + 2K(x, x_2) - K(x, x_3) + K(x, x_4)$$



How do we find good α_i?

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

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

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

Activity 21, problem 2

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

$$\widehat{y} = \underbrace{\boldsymbol{\phi}(\boldsymbol{x})^T \boldsymbol{\Phi}^T}_{[K(\boldsymbol{x}, \boldsymbol{x}_1) \quad K(\boldsymbol{x}, \boldsymbol{x}_2) \quad \cdots]} \overbrace{\underbrace{(\boldsymbol{\Phi\Phi}^T + \lambda \boldsymbol{I})^{-1} \boldsymbol{y}}_{\boldsymbol{\alpha}}}^{\boldsymbol{w}^*}$$

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

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

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

Danger! $\longrightarrow$ Overfitting

Use cross-validation to choose params.