

An aerial photograph of a city, likely Madison, Wisconsin, taken from a high vantage point looking down at the city and the surrounding water. The sun is setting behind a hill in the background, casting a warm, golden glow over the entire scene. The city is densely packed with buildings, and the water is filled with numerous sailboats. A large, semi-transparent rectangular box is overlaid on the center of the image, containing the course title in large, bold, black text.

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: $\hat{y} = \text{sign}(\mathbf{x}^T \mathbf{w})$

Linear regression: $\hat{y} = \mathbf{x}^T \mathbf{w}$

Kernels (in Machine Learning)

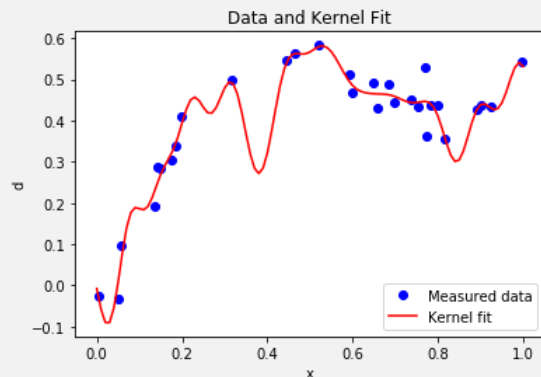
\mathbf{w} depends on $\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \dots$

Kernel

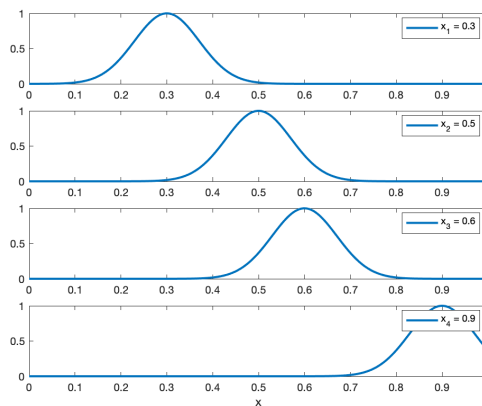
$$\hat{y} = \phi(\mathbf{x})^T \mathbf{w} \rightarrow \hat{y} = \sum_i \alpha_i K(\mathbf{x}, \mathbf{x}_i)$$

High dimensional

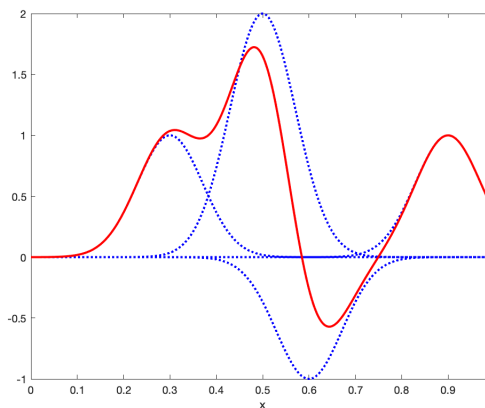
weighted sum of similarities between feature vector and each training point



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



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



How do we find good α_i ?

start by finding \mathbf{w} using ridge regression

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

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

↓ Activity 21, problem 2

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

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

$$\hat{y} = \phi(\mathbf{x})^T \mathbf{w}^* \rightarrow \hat{y} = \sum_i \alpha_i K(\mathbf{x}, \mathbf{x}_i)$$

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

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

Danger! → Overfitting

Use cross-validation to choose params.