

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, creating a warm, golden glow over the entire scene. The city is built on a peninsula, with numerous buildings and streets visible. The water is a deep blue, and several sailboats are scattered across it. A large, semi-transparent rectangular box is overlaid on the center of the image, containing the course title in large, black, sans-serif font.

# CS/ECE/ME 532 Matrix Methods in Machine Learning

*Welcome!*



# Announcements



Mon	Tue	Wed	Thu	Fri
			04/25 <b>Act. 22</b>	04/26
04/29	04/30 <b>Act. 23 (due 12/05)</b> <b>Act. 24 (due 12/07)</b>	05/01 <b>Assignment 9*</b>	05/02 <b>Act. 24 due</b> <b>Review</b>	05/03 <b>Assignment 10*</b>
05/06	05/07 <b>Assessment 4</b> <b>@ 5:35pm</b>			



**If you have a conflict, you must let me know TODAY!**  
**(fill out “Request for taking Assessment 4 at a different time”)**

\* We will drop the \*two\* lowest Assignments (instead of only one)



# Activity 22



Recall:

$w$  depends on  $x_1, y_1, x_2, y_2, \dots$

Kernel

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

High-dimensional feature vector

weighted sum of similarities between feature vector and each training point

$$w^* = \arg \min_w \|\Phi w - y\| + \lambda \|w\|^2$$

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



Activity 21, problem 2

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

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

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

where  $\Phi \Phi^T$  has  $i, j$  entry  $K(x_i, x_j)$

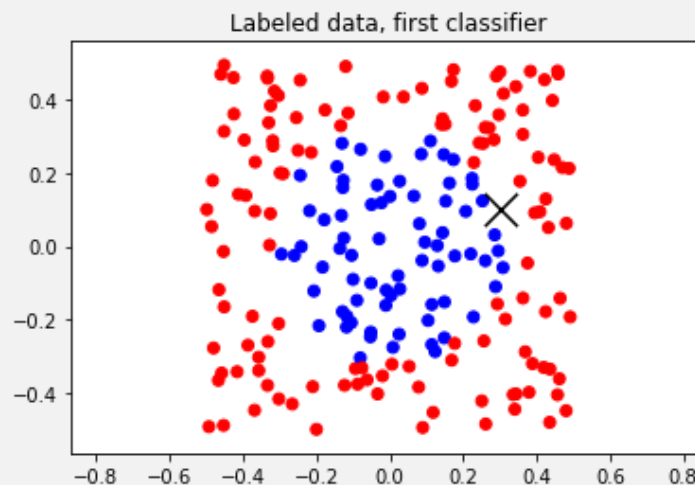
No need to compute  $\phi(\cdot)$  to compute  $K(\cdot, \cdot)$  or  $\hat{y}$ !

Some Insights:

$$\text{If } \lambda \uparrow, \quad w \downarrow \quad \text{and} \quad \hat{y}(x) \downarrow$$

$$\begin{array}{ccc} K(x, x_i) & & \hat{y}(x) \\ \text{Narrow} & \longrightarrow & \text{Rough} \\ \text{Broad} & \longrightarrow & \text{Smooth} \end{array}$$

Kernel classification



Kernel SVMs

$$\hat{y} = \text{sign} \left( \sum_i \alpha_i K(x, x_i) \right)$$

How do we find good  $\alpha_i$ ?

→ Iterative methods – subgradient descent

Support vectors have nonzero  $\alpha_i$

$$x = \begin{bmatrix} 0.3 \\ 0.1 \end{bmatrix}$$

How do we predict class of  $x$ ?

$$\hat{y} = \text{sign} \left( \sum_i \alpha_i K(x, x_i) \right)$$

$$K(x, x_i) = \exp \left( -\|x - x_i\|^2 \right)$$

$$\hat{y} = \text{sign} \left( \sum_i \alpha_i \exp \left( -\left\| \begin{bmatrix} 0.3 \\ 0.1 \end{bmatrix} - x_i \right\|^2 \right) \right)$$