

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 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



- Mid Course Feedback Survey
- This week:
  - ~~Last class: finish Unit 4~~
  - Today: start Unit 5
- Next week: spring break
- After spring break:
  - Tuesday: Assessment 3 review
  - Thursday: Assessment 3
- Assignment 7 is due the week after Assessment 3, but:
  - Q1 on the Eigendecomposition (attempt before the assessment)



# Activity 16

Hints for 1: video lecture 2.8

$$\mathbf{w}^* = \mathbf{V}\Sigma^{-1}\mathbf{U}^T\mathbf{y}$$

$\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w} = c$  define an ellipse  
with  $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$   
then columns of  $\mathbf{V}$  define axis



Before:

loss function  $\rightarrow$   $\min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2 + \lambda \|\mathbf{w}\|_2^2$   $\leftarrow$  regularizer

(This encourages solutions with small 2-norm)

closed form solution

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

What if we want some other regularizer?

(maybe only a few non-zero entries)

$$\min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2 + \lambda r(\mathbf{w})$$

Or different a loss function all together?

$$\min_{\mathbf{w}} f(\mathbf{w})$$

In general: no closed-form solution  $\rightarrow$  Iterative methods

## Gradient Descent

Main idea: use the gradient to head downhill

goal:  $\min_{\mathbf{w}} f(\mathbf{w})$   
for  $k = 1 \dots$   
 $\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \tau \nabla f(\mathbf{w})$   $\leftarrow$  step size

Gradient descent for least-squares:

goal:  $\min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2$   
for  $k = 1 \dots$   
 $\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \tau (\mathbf{X}^T \mathbf{X} \mathbf{w}^{(k)} - \mathbf{X}^T \mathbf{y})$

Convergence guaranteed if:

$$0 < \tau < \frac{2}{\sigma_1^2}$$

Steepest direction is  
normal to the contour!

