

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 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, bold, black text.

CS/ECE/ME 532 Matrix Methods in Machine Learning

Welcome!



Activity 18



Proximal Gradient Descent

Key idea: alternate gradient descent for LS with regularization

goal: $\min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda r(\mathbf{w}) \leftarrow \text{May not be differentiable!}$

set \mathbf{w}_0

for $k = 1 \dots$

$$\mathbf{z}^{(k)} = \mathbf{w}^{(k)} - \tau \mathbf{X}^T (\mathbf{X} \mathbf{w}^{(k)} - \mathbf{y})$$

Gradient Descent

$$\mathbf{w}^{(k+1)} = \arg \min_{\mathbf{w}} \|\mathbf{z}^{(k)} - \mathbf{w}\|_2^2 + \lambda \tau r(\mathbf{w})$$

Regularize

stay close to \mathbf{z} , but regularize

Least Absolute Shrinkage & Selection Operator

Regularized least squares with $r(\mathbf{w}) = \|\mathbf{w}\|_1$

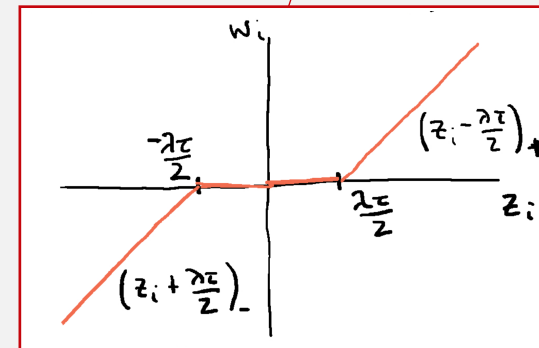
→ LASSO favors sparse solutions



Proximal Gradient Descent for LASSO:

$$\mathbf{w}^{(k+1)} = \arg \min_{\mathbf{w}} \sum (z_i^{(k)} - w_i)^2 + \lambda \tau |w_i|$$

$$\rightarrow w_i^{(k+1)} = (|z_i| - \lambda \tau / 2)_+ \text{sign}(z_i)$$



Proximal Gradient Descent for ridge regression

$$r(\mathbf{w}) = \|\mathbf{w}\|_2^2$$

$$\rightarrow \mathbf{w}^{(k+1)} = \frac{\mathbf{z}^{(k)}}{1 + \lambda \tau}$$

stay close to \mathbf{z} , but L2-shrink

