


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's buildings are visible on the left, and the water is filled with several 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!



Activity 17

$$\|g\|_2^2 = g^T g = \sum_{i=1}^M g_i^2$$


Gradient Descent

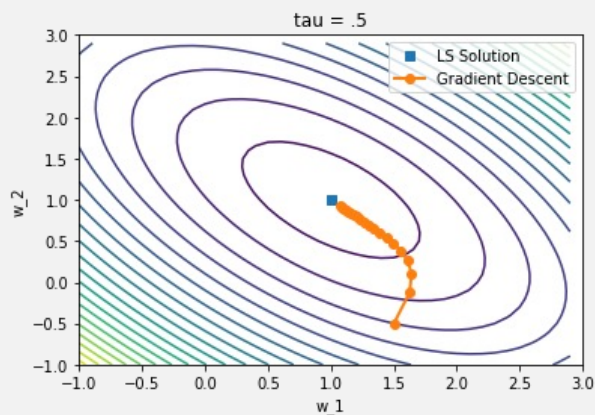
Main idea: use the gradient to head downhill

$$\begin{aligned} &\text{goal: } \min_{\mathbf{w}} f(\mathbf{w}) \\ &\text{for } k = 1 \dots \\ &\quad \mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \tau \nabla f(\mathbf{w}) \end{aligned}$$

step size

Gradient descent for least-squares:

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



Proximal Gradient Descent

Key idea: alternate gradient descent for LS with regularization

$$\begin{aligned} &\text{goal: } \min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda r(\mathbf{w}) \\ &\text{set } \mathbf{w}_0 \\ &\text{for } k = 1 \dots \\ &\quad \mathbf{z}^{(k)} = \mathbf{w}^{(k)} - \tau \mathbf{X}^T (\mathbf{X} \mathbf{w}^{(k)} - \mathbf{y}) \\ &\quad \mathbf{w}^{(k+1)} = \arg \min_{\mathbf{w}} \|\mathbf{z}^{(k)} - \mathbf{w}\|_2^2 + \lambda \tau r(\mathbf{w}) \end{aligned}$$

Gradient Descent

Regularization Step

stay close to \mathbf{z} , but regularize

For ridge regression:

$$\begin{aligned} r(\mathbf{w}) &= \|\mathbf{w}\|_2^2 \\ \Rightarrow \mathbf{w}^{(k+1)} &= \frac{\mathbf{z}^{(k)}}{1 + \lambda \tau} \end{aligned}$$

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

