CS/ECE/ME 532
Matrix Methods in Machine Learning

Welcome!

Activity 17

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

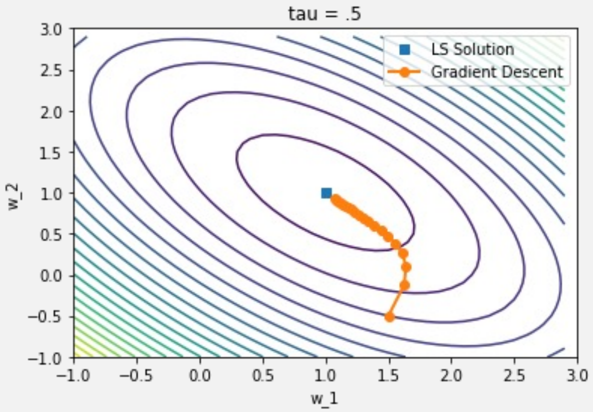
Gradient Descent

Main idea: use the gradient to head downhill

$$\begin{aligned}
&\text{goal: } \min_{\boldsymbol{w}} f(\boldsymbol{w}) \\
&\text{for } k = 1\ldots \\
&\quad \boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} - \underbrace{\tau}_{\text{step size}}\nabla f(\boldsymbol{w})
\end{aligned}$$

Gradient descent for least-squares:

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



Proximal Gradient Descent

Key idea: alternate gradient descent for LS with regularization

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

Gradient Descent

Regularization Step

stay close to z, but regularize

For ridge regression:

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

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

stay close to z, but L2-shrink