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: $\min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2 + \lambda r(\boldsymbol{w})$ ← *May not be differentiable!*

set $\boldsymbol{w}_0$

for $k = 1 \ldots$

Gradient Descent

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

Regularize

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

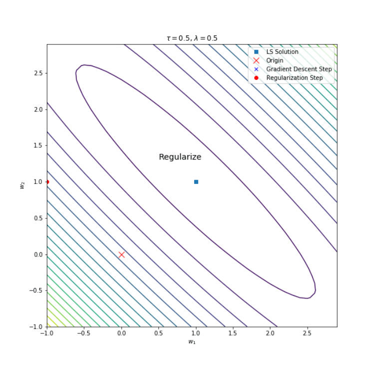
stay close to z, *but regularize*

Proximal Gradient Descent for ridge regression

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

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

stay close to z, but L2-shrink



Least Absolute Shrinkage & Selection Operator

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

→ *LASSO favors sparse solutions*

Proximal Gradient Descent for LASSO:

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

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