

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

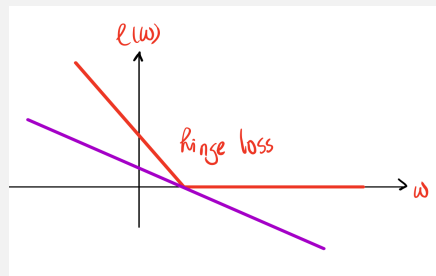
Welcome!



Activity 20



Sub-gradients

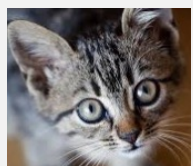


Hinge loss:
Convex
Non differentiable

Sub-gradient: any plane that lies below function.
any \mathbf{v} such that $\ell(\mathbf{w}) \geq \ell(\mathbf{w}_0) + (\mathbf{w} - \mathbf{w}_0)^T \mathbf{v}$

Classifying new data:

$\mathbf{x} =$



$$\hat{y} = \text{sign}(\mathbf{x}^T \mathbf{w})$$

if $\hat{y} = 1$ then dog

if $\hat{y} = -1$ then cat

$(\mathbf{x}_i, y_i), i = 1, \dots, \text{a million}$



Training a classifier:

a million

$$\min_{\mathbf{w}} \sum_{i=1}^{\text{a million}} (\mathbf{x}_i^T \mathbf{w} - y_i)^2$$

Problem: computing the loss is too slow.

Stochastic Gradient Descent

$$\min_{\mathbf{w}} \sum_{i=1}^{\text{a million}} \ell_i(\mathbf{w})$$

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \tau \nabla \ell(\mathbf{w}^k)$$

(Gradient Descent)

Main idea

Do gradient descent, but on a random subset of training examples at each iteration.

$$\mathbf{w}^{(1)} = \mathbf{w}^{(0)} - \tau \sum_{i=1}^{100} \nabla \ell_i(\mathbf{w}^{(0)})$$

"mini batch"

$$\mathbf{w}^{(2)} = \mathbf{w}^{(1)} - \tau \sum_{i=101}^{200} \nabla \ell_i(\mathbf{w}^{(1)})$$

Good for ML problems that involve lots of training data:

- Image/video classification and recognition
- ML translation
- Large scale prediction and regression tasks