9. Closing Comment
Closing Comment

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

(Optional) Equivalance of regularization to a Gaussian Prior on Weights

(Optional) Equivalance of regularization to a Gaussian Prior on Weights

The regularized linear regression can be interpreted from a probabilistic point of view. Suppose we are fitting a linear regression model with n data points (x_1, y_1),(x_2, y_2),...,(x_n, y_n). Let's assume the ground truth is that y is linearly related to x but we also observed some noise ϵ for y:

$$y_t = \theta \cdot x_t + \epsilon$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

Then the likelihood of our observed data is

$$\prod_{t=1}^{n} \mathcal{N}(y_t | \theta x_t, \sigma^2).$$

Now, if we impose a Gaussian prior $\mathcal{N}(\theta | 0, \lambda^{-1})$, the likelihood will change to

$$\prod_{t=1}^{n} \mathcal{N}(y_t | \theta x_t, \sigma^2) \mathcal{N}(\theta | 0, \lambda^{-1}).$$

Take the logarithim of the likelihood, we will end up with

$$\sum_{t=1}^{n} -\frac{1}{2\sigma^2}(y_t - \theta x_t)^2 - \frac{1}{2}\lambda\|\theta\|^2 + \text{constant}.$$

Try to derive this result by yourself. Can you conclude that maximizing this loglikelihood equivalent to minimizing the regularized loss in the linear regression? What does larger λ mean in this probabilistic interpretation? (Think of the error decomposition we discussed.)

Hide

Discussion

Hide Discussion

Topic: Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks):Lecture 5. Linear Regression / 9. Closing Comment

Add a Post

Show all posts by recent activity

Help clarify the notation in the optional problem	2
Great Lecture! Amazing. A difficult mathematical equation expressed in simple words. Great teacher!	13
[STAFF] What does larger λ mean in this probabilistic interpretation? (Think of the error decomposition we discussed.) How it is related to the error decomposition is unclear to me	2
Question regarding Guassian weights (Optional Section)	8
Why is minimizing the norm of theta increasing the error? Minimizing the norm of theta increases the error? or perhaps did I get it wrong from the lect...	2
Question about Gausian prior I want to know whether the first initialize of theta to be 0 caused the prior N(0,1/λ) or Not. If it...	2

The probabilistic or Bayesian interpretation is very interesting 1

That explains why we choose the mean square loss and quadratic regularization, which earlie...

Learn About Verified Certificates