Closing Comment

0:00 / 0:00 1.25x

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

< Previous   Next >