

<u>Unit 2 Nonlinear Classification</u>, <u>Linear regression, Collaborative</u>

Course > Filtering (2 weeks)

9. Closing Comment

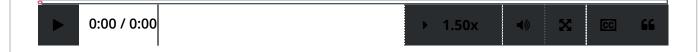
> Lecture 5. Linear Regression >

#### **Audit Access Expires May 11, 2020**

You lose all access to this course, including your progress, on May 11, 2020. Upgrade by Mar 25, 2020 to get unlimited access to the course as long as it exists on the site. **Upgrade now** 

# 9. Closing Comment Closing Comment





#### Video

Download video file

## **Transcripts**

<u>Download SubRip (.srt) file</u> Download Text (.txt) file

## (Optional) Equivalence of regularization to a Gaussian Prior on Weights (Optional) Equivalence 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  $\epsilon$  for y:

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

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

Then the likelihood of our observed data is

$$\prod_{t=1}^{n}\mathcal{N}\left(y_{t}| heta x_{t},\sigma^{2}
ight).$$

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

$$\prod_{t=1}^{n}\mathcal{N}\left(y_{t}| heta x_{t},\sigma^{2}
ight)\mathcal{N}\left( heta|0,\lambda^{-1}
ight).$$

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

$$\sum_{t=1}^n -rac{1}{2\sigma^2}(y_t- heta x_t)^2 -rac{1}{2}\lambda \| heta\|^2 + ext{constant}.$$

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  $\lambda$  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 teached	13 er!
[STAFF] What does larger λ mean in this probabilistic interpretation? (The error decomposition we discussed.) How it is related to the error decomposition is unclear to me	nink of the 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	2 n the lect
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	or Not. If it

9. Closing Comment | Lecture 5. Linear Regressio... https://courses.edx.org/courses/course-v1:MITx+...

The probabilistic or Bayesian interpretation is very interesting

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

**Learn About Verified Certificates** 

© All Rights Reserved