8. Regularization

Ridge Regression

0:00 / 0:00 1.50x

Video
[Download video file]

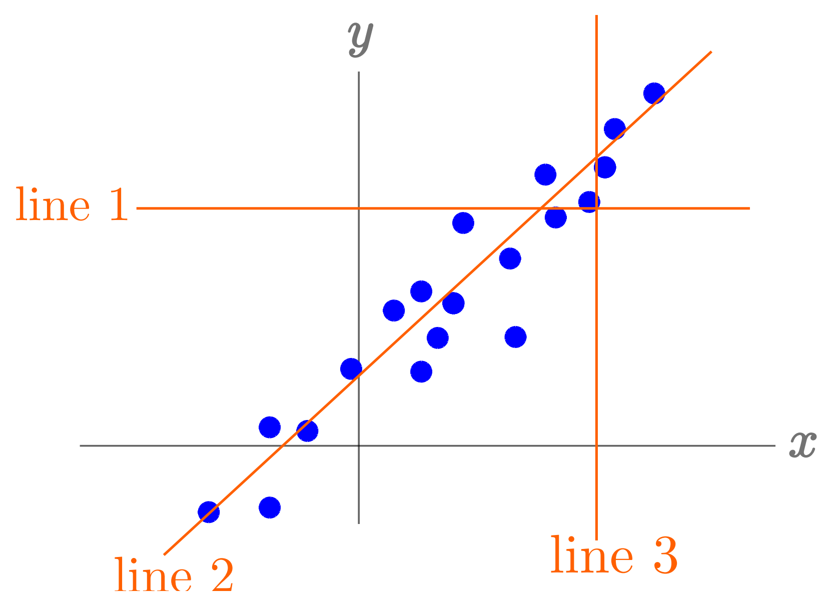
Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Regularization: extreme case 1

0/1 point (graded)
As in the video above, define the loss function

$$J_{n,\lambda}(\theta,\theta_0) = \frac{1}{n}\sum_{t=1}^{n}\frac{\left(y^{(t)} - \theta\cdot x^{(t)} - \theta_0\right)^2}{2} + \frac{\lambda}{2}\|\theta\|^2$$

where λ is the regularization factor.



In the figure above, the blue dots are the training examples. If we increase λ to ∞, where does $f(x) = \theta\cdot x + \theta_0$ converge to?

- line 1 ✔
- line 2
- line 3 ✖

Solution:

If we increase λ to ∞, minimizing J is equivalent to minimizing $||\theta||$. Thus θ will have to be a zero vector. Thus $f\left(x\right) = \theta \cdot x + \theta_0$ becomes $f\left(x\right) = \theta_0$, a horizontal line. Thus f converges to line 1.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

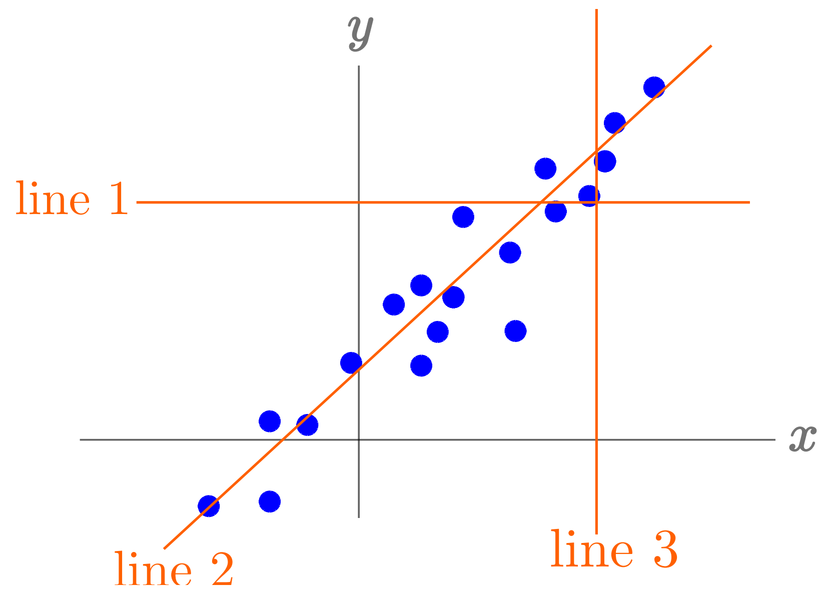
Regularization: Extreme case 2

1/1 point (graded)

As in the problem above,

$$J_{n,\lambda}\left(\theta, \theta_0\right) = \frac{1}{n} \sum_{t=1}^{n} \frac{\left(y^{(t)} - \theta \cdot x^{(t)} - \theta_0\right)^2}{2} + \frac{\lambda}{2} \left\|\theta\right\|^2$$

where λ is the regularization factor.



In the figure above, the blue dots are the training examples. If we decrease λ to 0, where does $f(x) = \theta \cdot x + \theta_0$ converge to?

- line 1
- line 2
- line 3

Solution:

If we decrease λ to zero, minimizing J is equivalent to minimizing $\frac{1}{n}\sum_{t=1}^{n}\frac{\left(y^{(t)}-\theta \cdot x^{(t)}-\theta_0\right)^2}{2}$, which is the "fit." Thus f converges to line 2.

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks):Lecture 5. Linear Regression / 8. Regularization

Add a Post

Show all posts by recent activity

- what is the eta*lambda*theta do in the update 3
- Why does the derivative of square norm of theta is 2*theta 3
  When calculating the gradient of the regularization term, the gradient/derivative of [lambda/...
- [Staff] Typo mistake with derivation on the board 5
  I think the professor forgot to include the summation sign in the derivation (when she took t...
- why theta_0 is not regularized? 3
  hi what's the reason that theta_0 is not included in the regularization component of the loss f...

Learn About Verified Certificates