MITx: 6.86x
Machine Learning with Python-From Linear Models to Deep Learning

8. Regularization

Ridge Regression

9:51 / 9:51 Speed 1.50x

this new reach

regression formula, apply gradient-based algorithm,

and to find the best theta.

In the homework, you will see-- in the exercise,

you will see how you can do the same kind of modification,

very straightforward modification,

to get a solution for this objective for the closed form

algorithm.

End of transcript. Skip to the start.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

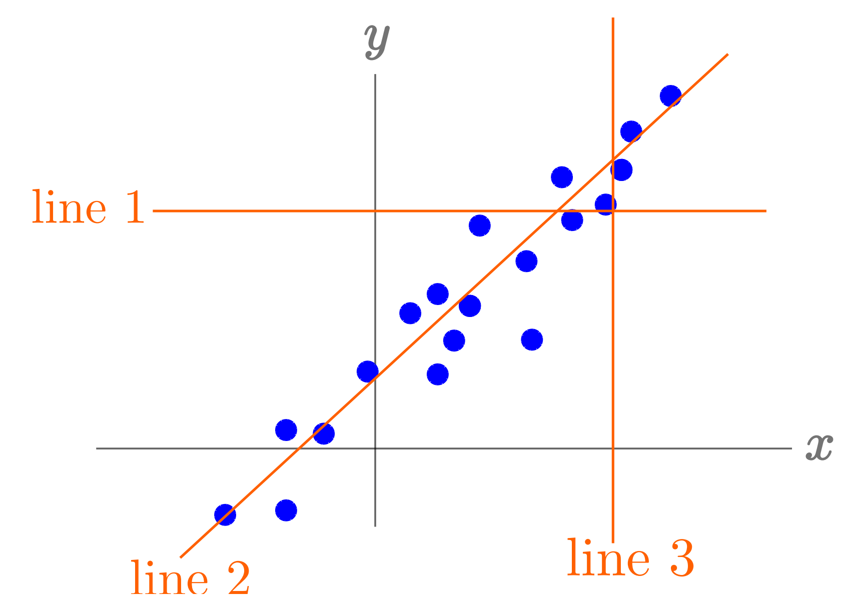
Regularization: extreme case 1

1/1 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 $f(x) = \theta \cdot x + \theta_0$ becomes $f(x) = \theta_0$, a horizontal line. Thus f converges to line 1.

Submit You have used 2 of 2 attempts

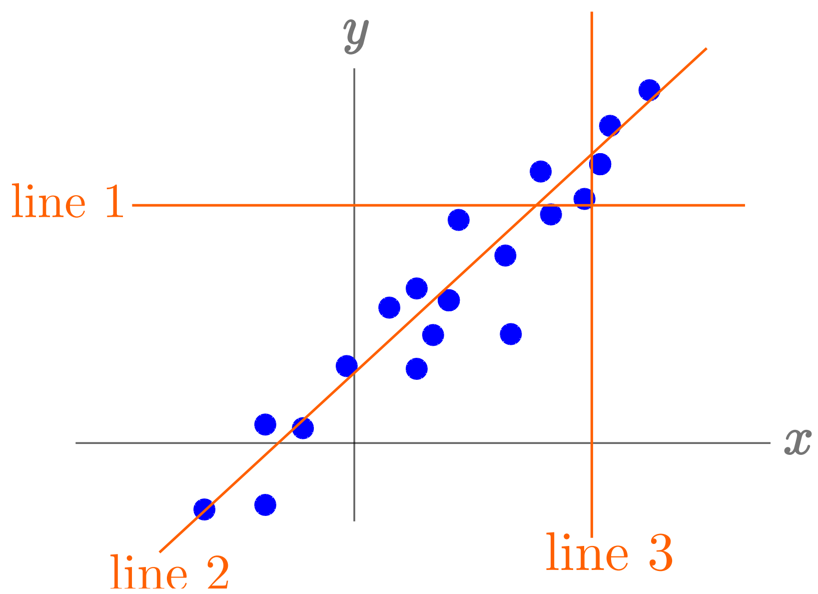
Answers are displayed within the problem

Regularization: Extreme case 2

1/1 point (graded)
As in the problem above,

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

where λ is the regularization factor.



In the figure above, the blue dots are the training examples. If we 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 ▼

? Why the term "regularization"?
What is "regular" about the modified objective function? Is there some deeper reason why the term "regularization" is used? For example how we distinguish perturbations as ... 1

Maybe a typo
00:00:17,320 --> 00:00:20,950
 And it's called 【reach regression】.

 138
 00:09:20,110 --> 00:09:27,340
 So at this point, you've seen how we can take ... 1

? Why adding regularization can help the model become more general?
I notice that adding regularization can make the parameter more close to zero. How I still confused about why the parameter more close to zero, the more general the model ... 2

? Regularization: Extreme case 2: the answer explanation is missing theta_0 3

Regularization: theta_0 left out of regularization, is this typical? 3

Regulatization only effects positive theta updates to prevent increasing theta values but not decreasing of theta value ?
Can someone please clarify, if regularization effects only when theta updates are positive and not when theta update is negative, that is it does not effect update step when t... 4

[Staff] Regularization: extreme case 1 and 2

Community TA

2

Learn About Verified Certificates