8. Regularization

Exercises due Mar 8, 2023 08:59 -03 Completed

Ridge Regression

0:00 / 0:00 1.0x

Video
Download video file

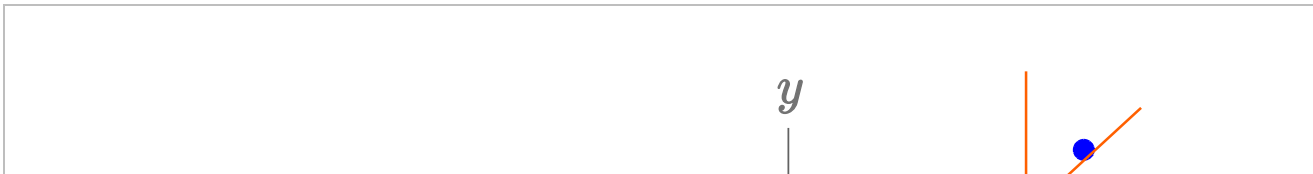
Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

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.



line 1

line 2

line 3

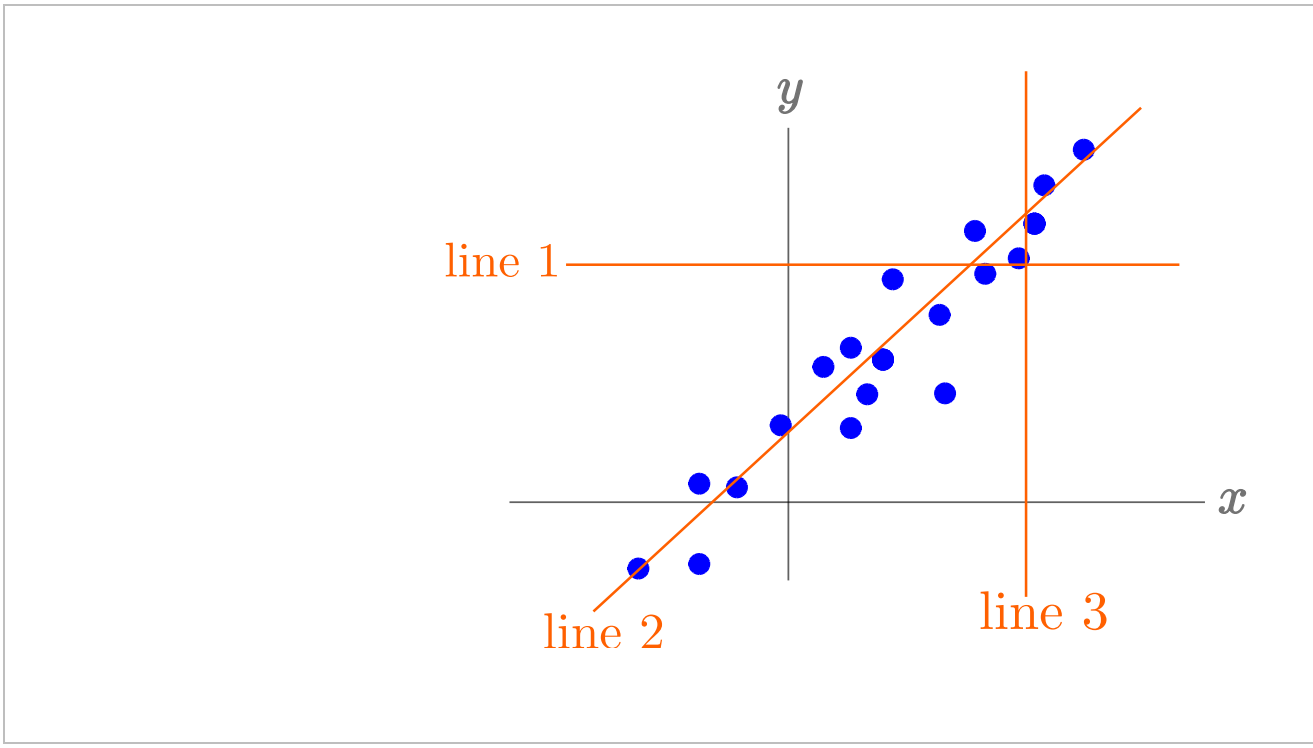
Submit You have used 2 of 2 attempts

Regularization: Extreme case 2

1/1 point (graded)

As in the problem above,

where is the regularization factor.



In the figure above, the blue dots are the training examples. If we decrease to , so (shorthand for the weights optimal with respect to and the depicted data) also chang does the predictor line (i.e., the graph of the function from s to

Show all posts

- Regularization with large data sets
  Let us assume we have a very large data set at hand, and a not-so-complex model. Do we still need to do re
- possible typo
  In part two, should the above question say "if we decrease lambda to negative infinity..."? I can't really make
- taking derivative of theta while it is a norm
- Found an error in the transcript
  ridge regression not *reach* regression. 0:17 :)
- mistake information in "Regularization: Extreme case 2"
  Regularization: Extreme case 2: "In the figure above, the blue dots are the training examples. If we decrease
- possible plotting issue!
  Shouldn't Line 1 have been drawn lower, so that its location corresponds to the average y values?
- Need intuition on the regularization term
  Why is the regularization term defined in terms of norm(theta)? For linear classifiers, it makes sense why reg
- 8:38 (1-eta*lambda)*theta expression
  I don't understand why "at every point of our average we're actually pushing thetas down" if the learning rate
- If regularization is high, the points are not adjusted as much as if regularization is low, the poi adjusted a lot?
  So how does this translate to the graph as shown above? Like if the points get adjusted, how will i know if it