





MITx 6.86x

**Machine Learning with Python-From Linear Models to Deep Learning**[Course](#)[Progress](#)[Dates](#)[Discussion](#)[Resources](#)[Course](#) / [Unit 2. Nonlinear Classification, Linear regression, Col...](#) / [Lecture 5.](#)[< Previous](#)

## 8. Regularization

Bookmark this page

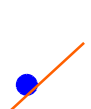
Exercises due Mar 8, 2023 08:59 -03 Completed

**Ridge Regression****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

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

where  $\lambda$  is the regularization factor. $y$ 



line 1



line 2



line 3



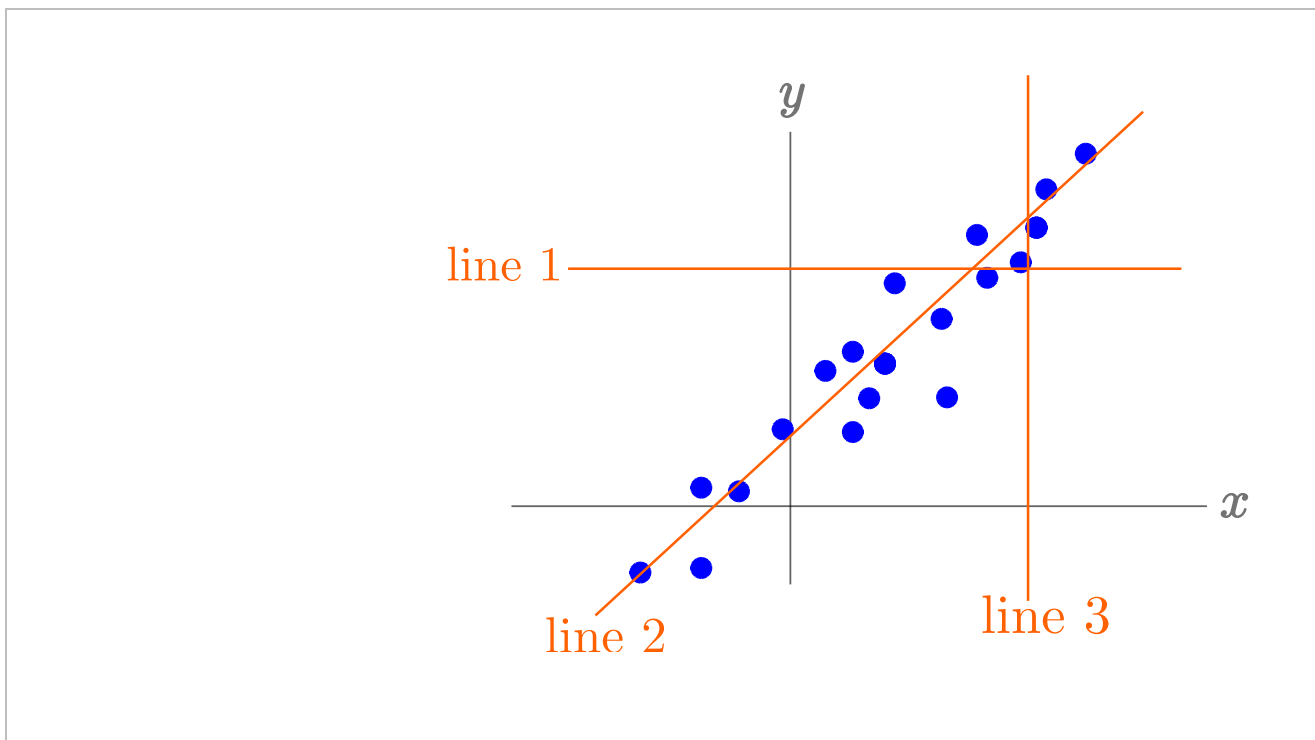
Submit

You have used 2 of 2 attempts

## Regularization: Extreme case 2

1/1 point (graded)

As in the problem above,

where  $\lambda$  is the regularization factor.

In the figure above, the blue dots are the training examples. If we decrease  $\lambda$  to 0, so  $\lambda \rightarrow 0$ ,  $w$  (shorthand for the weights optimal with respect to  $\lambda$  and the depicted data) also changes. As  $\lambda$  decreases, does the predictor line (i.e., the graph of the function  $f(x) = w_0 + w_1 x$ ) change from  $s$  to  $s'$ ?

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

## edX

[About](#)[Affiliates](#)[edX for Business](#)[Open edX](#)[Careers](#)[News](#)

## Legal

[Terms of Service & Honor Code](#)[Privacy Policy](#)



© 2023 edX LLC. All rights reserved.

深圳市恒宇博科技有限公司 [粤ICP备17044299号-2](#)