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

Help smitha_kannur

Course Progress Dates Discussion Resources

Course / Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks) / Homework 2

< Previous	✎	✎	✎	✎ ✓	✎	Next >

5. Linear Regression and Regularization

Bookmark this page

Homework due Oct 14, 2020 05:29 IST *Past Due*

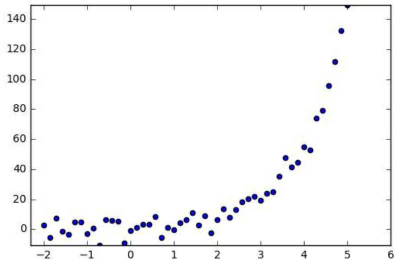
In this question, we will investigate the fitting of linear regression.

---

5. (a)

2/2 points (graded)

For each of the datasets below, provide a simple feature mapping ϕ such that the transformed data $\left(\phi\left(x^{(i)}\right), y^{(i)}\right)$ would be well modeled by linear regression.



Which feature mapping ϕ is appropriate for the above model?

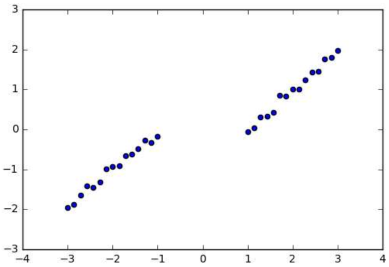
● $\exp(x)$

○ $\log(x)$

○ x^2

○ $\sqrt{x}$

✓



Which feature mapping ϕ is appropriate for the above model?

○ $\phi(x) = x + \text{sign}(x)$

● $\phi(x) = x - \text{sign}(x)$

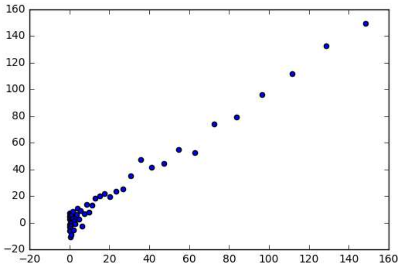
○ $\phi(x) = x \cdot \text{sign}(x)$

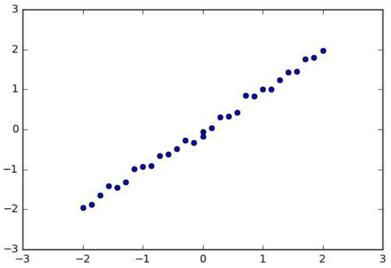
○ $\phi(x) = x / \text{sign}(x)$

✓

Solution:

- In both figures the data seem to follow a non-linear pattern so they would not be fit well by a linear model.
- We can, however, use a non-linear transformation $\phi\left(x\right)$ so that, in the new feature space, a linear model produces a good fit.
- In the 1st plot, the data seem to roughly follow $y = e^x$, so an exponential transformation, $\phi\left(x\right) = e^x$, would yield $\left(\phi\left(x^{(i)}\right), y^{(i)}\right)$ that could be fit well by linear regression.
- In the 2nd plot, the observations appear to be generated by the discontinuous function $y = x - \text{sign}\left(x\right)$ (where $\text{sign}\left(x\right) = x/|x|$), so if we let $\phi\left(x\right) = x - \text{sign}\left(x\right)$, an observation $y^{(i)}$ should be more easily modeled by a linear function of $\phi\left(x^{(i)}\right)$, which will be found by linear regression.
- The results of the transformations are plotted below.





Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

5. (b)

2 points possible (graded)

Consider fitting a ℓ_2-regularized linear regression model to data $\left(x^{(1)}, y^{(1)}\right), \ldots, \left(x^{(n)}, y^{(n)}\right)$ where $x^{(t)}, y^{(t)} \in \mathbb{R}$ are scalar values for each $t = 1, \ldots, n$. To fit the parameters of this model, one solves

$$\min_{\theta \in \mathbb{R},\ \theta_0 \in \mathbb{R}} L\left(\theta, \theta_0\right)$$

where

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

Here $\lambda \geq 0$ is a pre-specified fixed constant, so your solutions below should be expressed as functions of λ and the data. This model is typically referred to as **ridge regression** .

Write down an expression for the gradient of the above objective function in terms of θ.

Important: If needed, please enter $\sum_{t=1}^{n}\left(\ldots\right)$ as a function `sum_t(...)`, including the parentheses. Enter $x^{(t)}$ and $y^{(t)}$ as `x^{t}` and `y^{t}`, respectively.

$\frac{\partial L}{\partial \theta} =$

Answer: 2*lambda*theta - 2*sum_t((y^{t} - theta*x^{t} - theta_0)*x^{t})

Write down an expression for the gradient of the above objective function in terms of θ_0.

$\frac{\partial L}{\partial \theta_0} =$ **Answer:** -2*sum_t(y^{t} - theta*x^{t} - theta_0)

STANDARD NOTATION

Solution:

- The gradient is a two-dimensional vector $\nabla L = \left[\frac{\partial L}{\partial \theta_0}, \frac{\partial L}{\partial \theta}\right]$, where
- $\frac{\partial L}{\partial \theta_0} = -2\sum_{t=1}^{n}\left(y^{(t)} - \theta x^{(t)} - \theta_0\right)$
- $\frac{\partial L}{\partial \theta} = 2\lambda\theta - 2\sum_{t=1}^{n}\left(y^{(t)} - \theta x^{(t)} - \theta_0\right)x^{(t)}$

Submit You have used 0 of 5 attempts

Answers are displayed within the problem

5. (c)

2 points possible (graded)

Find the closed form expression for θ_0 and θ which solves the ridge regression minimization above.

Assume θ is fixed, write down an expression for the optimal $\hat{\theta}_0$ in terms of $\theta, x^{(t)}, y^{(t)}, n$.

Important: If needed, please enter $\sum_{t=1}^{n}(\ldots)$ as a function `sum_t(...)`, including the parentheses. Enter $x^{(t)}$ and $y^{(t)}$ as `x^{t}` and `y^{t}`, respectively.

$\hat{\theta}_0 =$ **Answer:** 1/n * sum_t(y^{t} - theta*x^{t})

Write down an expression for the optimal $\hat{\theta}$. To simplify your expression, use $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x^{(t)}$. Your answer should be in terms of $x^{(t)}, y^{(t)}, \lambda$ and $\bar{x}$ **only**.

Important: If needed, please enter $\sum_{t=1}^{n}(\ldots)$ as a function `sum_t(...)`, including the parentheses. Enter $x^{(t)}$ and $y^{(t)}$ as `x^{t}` and `y^{t}`, respectively. Enter $\bar{x}$ as `barx`.

$\hat{\theta} =$

Answer: (sum_t((x^{t} - barx)*y^{t})) / (lambda + sum_t(x^{t} * (x^{t} - barx)))

Now after the optimal $\hat{\theta}$ is obtained, you can use it to compute the optimal $\hat{\theta}_0$

Solution:

To find the θ, θ_0 which minimize L, we note that because this objective function is convex, any point where $\nabla L(\theta_0, \theta) = 0$ is a global minimum. Thus, we set the gradient equal to zero and solve for θ, θ_0 to find the minimizers:

$$\frac{\partial}{\partial \theta_0} = -2\sum_{t=1}^{n}\left(y^{(t)} - \theta x^{(t)} - \theta_0\right) = -2\sum_{t=1}^{n}\left(y^{(t)} - \theta x^{(t)}\right) + 2\sum_{t=1}^{n}\theta_0 = 0$$

$$\implies \quad -2n\theta_0 = -2\sum_{t=1}^{n}\left(y^{(t)} - \theta x^{(t)}\right) \quad \implies \quad \theta_0 = \frac{1}{n}\sum_{t=1}^{n}\left(y^{(t)} - \theta x^{(t)}\right)$$

$$\frac{\partial}{\partial \theta} = 2\lambda\theta - 2\sum_{t=1}^{n}\left(y^{(t)} - \theta x^{(t)} - \theta_0\right)x^{(t)}$$

$$= 2\lambda\theta - 2\sum_{t=1}^{n}\left(y^{(t)} - \theta x^{(t)} - \left[\frac{1}{n}\sum_{s=1}^{n}\left(y^{(s)} - \theta x^{(s)}\right)\right]\right)\cdot x^{(t)} = 0$$

$$\implies \lambda\theta - \sum_{t=1}^{n} x^{(t)}y^{(t)} + \theta\sum_{t=1}^{n} {x^{(t)}}^{2} + \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\left(y^{(s)} - \theta x^{(s)}\right)x^{(t)} = 0$$

$$\implies \lambda\theta - \sum_{t=1}^{n} x^{(t)}y^{(t)} + \theta\sum_{t=1}^{n} {x^{(t)}}^{2} + \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n} y^{(s)}x^{(t)} - \frac{1}{n}\theta\sum_{t=1}^{n}\sum_{s=1}^{n} x^{(s)}x^{(t)} = 0$$

$$\implies \hat{\theta} = \frac{\sum_{t=1}^{n} x^{(t)}y^{(t)} - \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n} y^{(s)}x^{(t)}}{\lambda + \sum_{t=1}^{n} {x^{(t)}}^{2} - \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n} x^{(s)}x^{(t)}} \quad \text{is the value of } \theta \text{ which minimizes } L\left(\theta_0, \theta\right).$$

Note that if we define $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x^{(t)}$, then we can rewrite the above expression in a nicer form:

$$\hat{\theta} = \frac{\sum_{t=1}^{n}\left(x^{(t)} - \bar{x}\right)y^{(t)}}{\lambda + \sum_{t=1}^{n} x^{(t)}\left(x^{(t)} - \bar{x}\right)}$$

In other words, adding an unpenalized bias is equivalent to training on a centered dataset.

Finally, we can plug this value of $\hat{\theta}$ back into expression $\hat{\theta}_0 = \frac{1}{n}\sum_{t=1}^{n}\left(y^{(t)} - \theta x^{(t)}\right)$ to find the corresponding $\hat{\theta}_0$ which together with $\hat{\theta}$ minimizes L.

Submit

You have used 0 of 5 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

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

Add a Post

Show all posts

by recent activity

Please help with the math side of the course. 6
I need help with the math side of the course that Prof.Jaakkola is teaching. help me with where to start with all the prerequisites nee...

? 5c answer explanation 1

keep having this message: Invalid Input: \'barx\' not permitted in answer as a variable 3
I have put this formula which seems right to me but keep having the above "barx" error message. Any idea? (sum_t(y^{t}*x^{t}-barx*...

If x is element of R2, then these derived equations wont hold. Is that right? 3
If x is element of R2, then these derived equations wont hold. Is that right?. Especially the 5c_2

[STAFF] the grader did not accept my right answer! 2
If i am not misleading y answer to the theta was as good as the suggested solution (sum_t(y^{t}*x^{t})-barx*sum_t(y^{t}))/(lambda+s...

? Any hint on 5(b)? 6
I'm trying to do a partial derivative of the expression, but I'm only geting the wrong answer

? Invalid input error 2
Im getting an invalid input error for questions 5b and 5c, even though the notations as well as answer are correct. Unable to submit....

? What does L in question 5(b) means? 2
Does it means Loss? it's a loss function?

? staff/someone please help 2
Hi, I keep getting the message > Invalid Input: \'barx\' not permitted in answer as a variable when I input my answer for the second p...

Hint - 5b & 5c - remember x^{t} is scalar
i assumed x^{t} is vector and struggled for 5c2
1

5c -part 2: I am still very confused
In trying to answer 5-c i have tried 2 interpretations for the regularized criterion L(theta,theta_0) given in question 5-b : A. the regular...
2

☑ a.)
Hello, By intuition, if the data has a particular shape, it can be made into a linear shape by using the inverse of given function. That d...
7

? Problem with progress update for homework 2
After having completed homework 2, I checked progress tab and found my grade is just 88% (22/25). Even I only missed to answer c...
5

Invalid Input

‹ Previous

Next ›

edX

About
Affiliates
edX for Business
Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

Blog
Contact Us
Help Center
Media Kit
Donate
