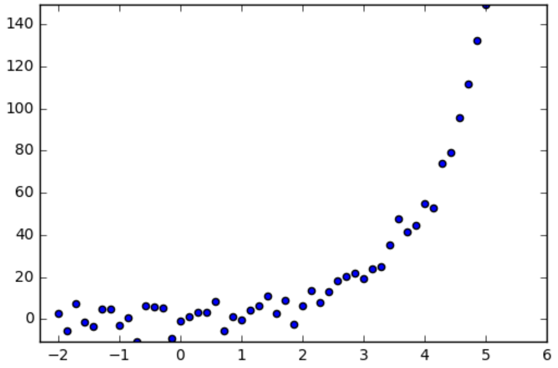
5. Linear Regression and Regularization

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.



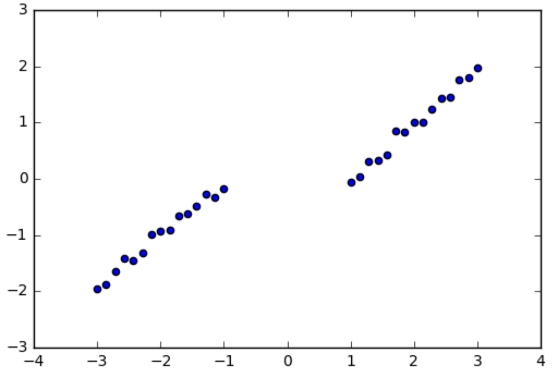
Which feature mapping ϕ is appropriate for the above model?

● $\exp(x)$ ✓

○ $\log(x)$

○ 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:

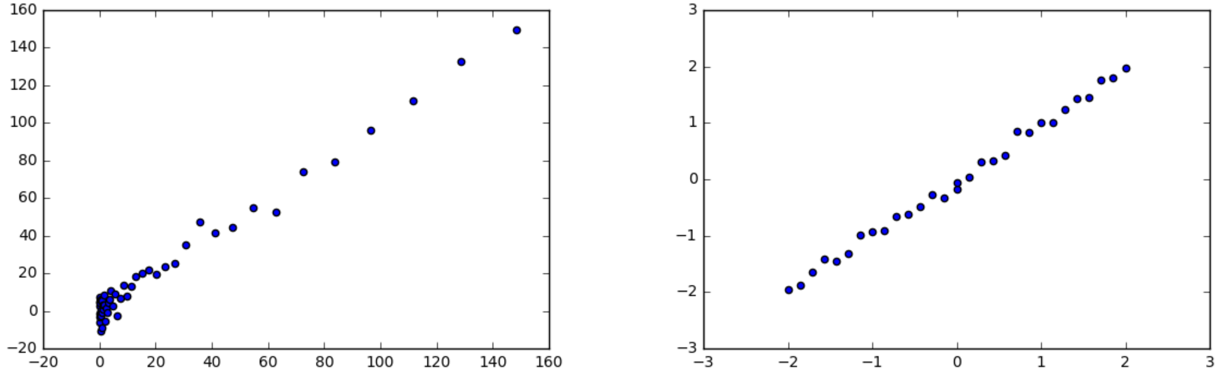
- In both figures the data seem to follow a non-linear pattern so they would not be fit well by a linear model.

- We can, however, use a non-linear transformation $\phi(x)$ 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, $\phi(x) = e^x$, would yield $(\phi(x^{(i)}), y^{(i)})$ that could be fit well by linear regression.

- In the 2nd plot, the observations appear to be generated by the discontinuous function $y = x - \text{sign}(x)$ (where $\text{sign}(x) = x/|x|$), so if we let $\phi(x) = x - \text{sign}(x)$, an observation $y^{(i)}$ should be more easily modeled by a linear function of $\phi(x^{(i)})$, 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)

2.0/2 points (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.

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

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.

$\frac{\partial L}{\partial \theta_0} =$ -2*sum_t(y^{t}-theta*x^{t}-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 1 of 5 attempts

Answers are displayed within the problem

---

5. (c)

2.0/2 points (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.

$\hat{\theta}_0 =$ (sum_t(y^{t}-theta*x^{t}))/n ✓ **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 $x^{(t)}$, $y^{(t)}$, λ 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`.

$\hat{\theta} =$ (sum_t(y^{t}*x^{t})-barx*sum_t(y^{t})) / (lambda + sum_t(✓

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:

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

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

$$\begin{aligned}\frac{\partial}{\partial\theta} &= 2\lambda\theta - 2\sum_{t=1}^{n}\left(y^{(t)} - \theta x^{(t)} - \theta_0\right)x^{(t)} \\ &= 2\lambda\theta - 2\sum_{t=1}^{n}\left(y^{(t)} - \theta x^{(t)} - \left[\frac{1}{n}\sum_{s=1}^{n}\left(y^{(s)} - \theta x^{(s)}\right)\right]\right)\cdot x^{(t)} = 0\end{aligned}$$

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

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

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

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 1 of 5 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks):Homework 3 / 5. Linear Regression and Regularization

Add a Post

Show all posts ▼ by recent activity ▼

? [STAFF] 5c correct solution (part 2) 3
I got theta_hat correct on my second-to-last attempt, but the grader marked it as incorrect. I tried to simplify my expression in my last attempt, b...

☑ [Staff] Right form of solution in 5c 6

💬 [Staff] 5(c) for theta_0 is not correct 4
Sorry I am not sure what is wrong. I am using the same equation which was correct from 5(b), and putting the expression for theta_0 equals to 0....

💬 Equivalent expression in denominator of 5. C part 2 2

💬 n not permitted in answer 7 new_
I just dont understand this. Why go through this trouble? I mean does this really add to knowledge? does it have any teaching value? I have been ...

☑ 5c2 Problem 2
I'm not sure I understand how I parse x^2 from the sum to extract theta. Could someone elaborate on that pls?

Q 5c : strange parsing error when output looks legit on screen 3
Hello When I try to submit I see the following error for answer for theta: > Invalid Input: Could not parse \'(lambda + sum_t***** as a formula I s...

☑ Unpenalized bias 2
What exactly is meant by "adding an unpenalized bias" in the solution to 5(c)? I see how the answer implies the dataset is centered, but I don't un...

☑ Staff - Question 5C part 2 7 new_ 15
Staff, Could you please take a look at my answer in c part 2? Submission 2 and 3? I believe I one of those is the correct answer. Thank you

☑ To staff: Could you please check my 5.C (part 2)? 4
Hi! I just cant figure out what Im doing wrong and I seem to get the same result every time. Could you help me?

☑ 5. (c) 34
I must have done something wrong in this exercise. Since the goal is minimizing the function, I started reasoning from the derivatives computed ...

☑ 5b - theta 9
I managed to get the derivative in relation to theta_0 right (the second box), but I am having problems with the theta (first box). I've taken the der...

☑ 5c. How to isolate theta 3

© All Rights Reserved