Introduction to Machine Learning

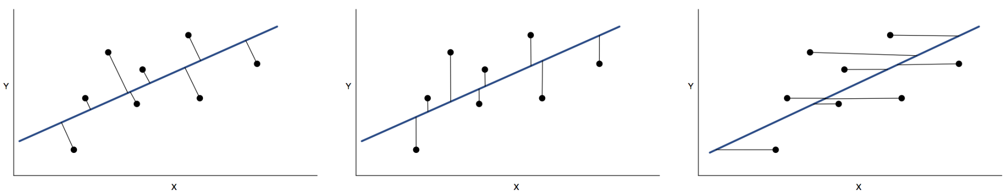
Due: Monday, September 11, 2023 at 09:00 AM

Before working through these questions, please study the lecture notes on Regression

1) Intro to linear regression

1.1) §

Squared error is frequently used as a loss function for regression. Which of the following pictures illustrates the squared loss function? Assume that the dark **blue** line is described by θ, θ_0, the black dots are the , data, and the light lines indicate the errors we are measuring.



A B C

Select the picture which best illustrates the squared loss function: B

100.00%

You have 1 submission remaining.

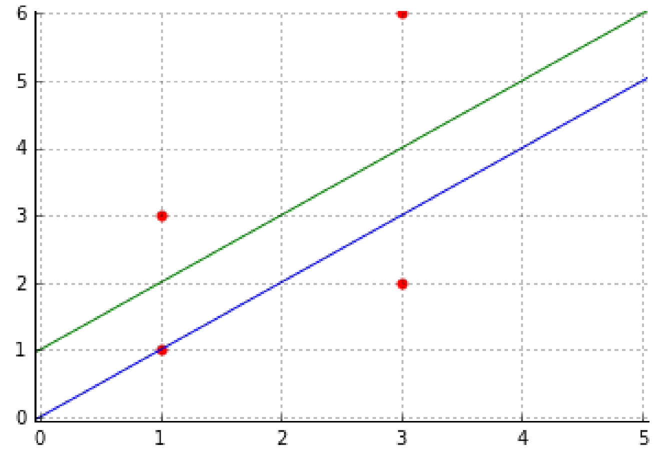
Solution: B

Explanation:

Squared loss measures the squared distance between the actual and the predicted θ θ_0. Therefore, the picture should depict distances in only , which corresponds to B.

1.2)

Consider the data set and regression lines in the plot below.



- The equation of the blue (lower) line is:
- The equation of the green (upper) line is:
- The data points (in x, y pairs) are: , , , , , , ,

What is the squared error of each of the points with respect to the **blue** line?

Provide a Python list of four numbers (in the order of the points given above).

[4,0,1,9]

100.00%

You have infinitely many submissions remaining.

Solution: `[4, 0, 1, 9]`

Explanation:

Given the data points and regression line , we calculate the squared error of each point:
At , the regression predicts , so the first two errors are and 0.
At , the regression predicts , so the last two errors are and .

1.3)

Basic linear regression seeks to minimize the mean squared error over all training points:

$$\theta, \theta_0 \quad \theta \;\; \theta_0$$

That means that the gradient (with respect to θ and θ_0) of the mean squared error regression criterion has the form of a sum over contributions from individual points. So,

$$\theta \quad \theta \;\; \theta_0$$

and

$\theta_0 \quad \theta \ \theta_0$

In the following questions, ignore the factor of and consider just the terms inside the sum.

What is the contribution from each point to the gradient of the objective with respect to the parameters θ and θ_0 of the **blue** (lower) line?

(Hint: re-express the above equation in terms of $\theta \ \theta_0$, the predicted value.)

Provide a list of four pairs of numbers (as tuples of θ, θ_0, in the order of the points given above).

[(-4, -4), (0, 0), (6,2), (-18, -6)]

100.00%

You have infinitely many submissions remaining.

Solution: `[(-4, -4), (0, 0), (6, 2), (-18, -6)]`

Explanation:

We should note that $\theta \ \theta_0$ is the predicted for the point . Thus, we can simplify the gradient formula to , . Also note that conceptually is the error described by θ, θ_0 on the point , . For the blue line, .
At , , , so the gradient is , , .
At , , , so the gradient is 0 , 0 $0, 0$.
At , , , so the gradient is , , .
At , , , so the gradient is , , .

1.4)

What is the squared error of each of the points with respect to the **green** line?

Provide a list of four numbers (in the order of the points given above).

[1, 1, 4, 4]

100.00%

You have infinitely many submissions remaining.

Solution: `[1, 1, 4, 4]`

Explanation:

Given data points , , , , , , , and green regression , we calculate the squared error of each point.

At , the regression predicts , so the first two errors are and .

At , the regression predicts , so the last two errors are and .

1.5)

What is the contribution from each point to the gradient of the objective with respect to the parameters of the green line?

Provide a list of four pairs of numbers (as tuples, in the order of the points given above).

[(-2,-2), (2,2), (12,4), (-12,-4)]

100.00%

You have infinitely many submissions remaining.

Solution: `[(-2, -2), (2, 2), (12, 4), (-12, -4)]`

Explanation:

We should note that θ θ_0 is the predicted for the point . Thus, we can simplify the gradient formula to , . Also note that conceptually is the error described by θ, θ_0 on the point , . For the green line, .

At , , , so the gradient is , , .

At , , , so the gradient is , , .

At , , , so the gradient is , , .

At , , , so the gradient is , , .

1.6)

Mark all of the following that are true:

- [ ] The blue line minimizes mean squared error.
- [x] The green line minimizes mean squared error.
- [ ] The mean squared error from all the points to the blue line is 0.
- [ ] The mean squared error from all the points to the green line is 0.
- [ ] The sum of the gradient contributions in all dimensions from all the points for the blue line is 0.
- [x] The sum of the gradient contributions in all dimensions from all the points for the green line is 0.
- [ ] Neither line minimizes mean squared error.
- [ ] It is impossible to minimize mean squared error.
- [ ] Both lines minimize mean squared error.

100.00%

You have infinitely many submissions remaining.

Solution:

- [ ] The blue line minimizes mean squared error.
- [x] The green line minimizes mean squared error.
- [ ] The mean squared error from all the points to the blue line is 0.
- [ ] The mean squared error from all the points to the green line is 0.
- [ ] The sum of the gradient contributions in all dimensions from all the points for the blue line is 0.
- [x] The sum of the gradient contributions in all dimensions from all the points for the green line is 0.
- [ ] Neither line minimizes mean squared error.
- [ ] It is impossible to minimize mean squared error.
- [ ] Both lines minimize mean squared error.

Explanation:

The blue line's mean squared error (MSE) is 14, while the green line's MSE is 10, so (1) is false and (2) is true. Neither (3) nor (4) are true.
The blue line's sum of gradient contributions is , , and the green line's is $0, 0$, so (5) is false and (6) is true.
The green line has a zero gradient, so it minimizes the mean squared error. It follows that (7), (8), and (9) are false.

2) Ridge regression

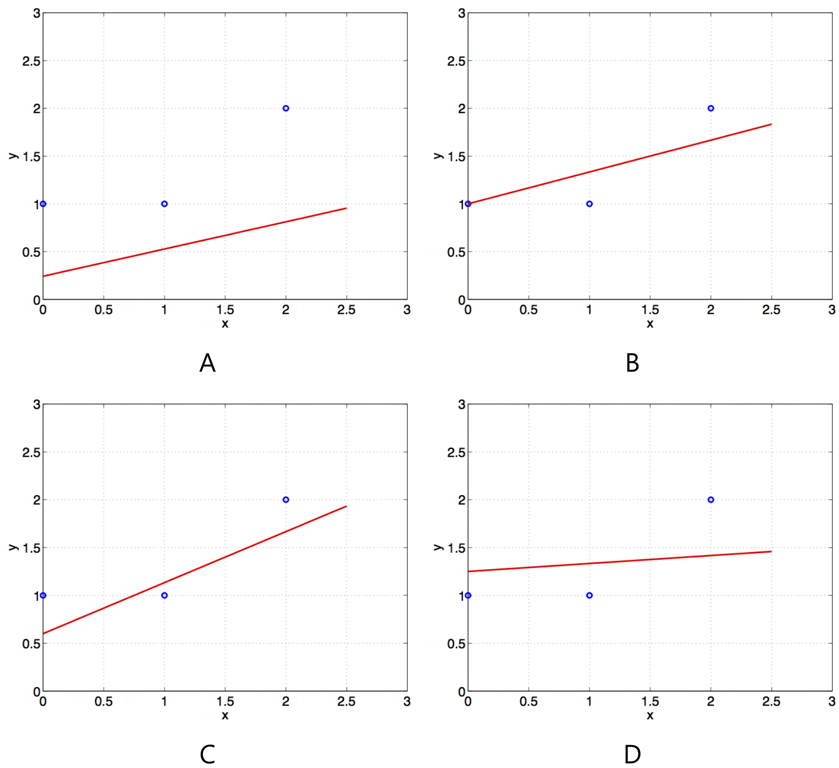
Consider a one-dimensional ordinary least squares regression problem, i.e., If we use the squared-norm regularizer on all parameters (i.e. both θ and θ_0), we get the canonical *ridge regression* objective:

$$\theta, \theta_0 \qquad , \theta, \theta_0 \;\; \theta \;\; \theta_0$$

where is the usual squared loss.

In practice, we do not have to strictly follow the canonical form of regularizing $\theta \;\; \theta_0$ Instead, we may add regularization penalty on either just θ, or on both θ and θ_0 We will consider the effect of these various choices of regularization.

The figures below plot linear regression results on the basis of only three data points , , , , , . We used various types of regularization to obtain the plots (see below) but got confused about which plot corresponds to which regularization method. Please assign each plot to one (and only one) of the following regularization methods.



2.1)

θ θ_0 θ where is plot: B

100.00%

You have 0 submissions remaining.

Solution: B

Explanation:

The regularization term θ only penalizes the weights but not the offset θ_0, so the slope of the line becomes closer to 0. When is small, the regression is nudged to be slightly more horizontal, but the -intercept does not move downwards (B).

2.2)

θ θ_0 θ where 0 is plot: D

100.00%

You have 1 submission remaining.

Solution: D

Explanation:

The regularization term θ only penalizes the weights but not the offset θ_0, so the slope of the line is more strongly nudged to become closer to 0. When is large, the regression approaches horizontal, but the -intercept does not approach 0 (D).

2.3)

θ θ_0 θ θ_0 where is plot: C

100.00%

You have 1 submission remaining.

Solution: C

Explanation:

The regularization term θ θ_0 penalizes both weights and bias. When is small, the line slightly approaches the axis, so the -intercept dips slightly (C).

2.4)

θ θ_0 θ θ_0 where 0 is plot: A

100.00%

You have 2 submissions remaining.

Solution: A

Explanation:

The regularization term θ θ_0 penalizes both weights and bias. When is large, the entire line approaches the axis, which has 0 slope and 0 -intercept (A).

Survey

(The form below is to help us improve/calibrate for future assignments; submission is encouraged but not required. Thanks!)

How did you feel about the **length** of this exercise?

- Too long.
- About right.
- Too short.

How did you feel about the **difficulty** of this exercise?

- Too hard. We should tone it down.
- About right.
- Too easy. I want more challenge.

Do you have any feedback or comments about any questions in this exercise? Anything else you want us to know?

Submit