Alexander Guðmundsson
alg35@hi.is

REI602M **Quiz 15.1.2021**

Hand in on Gradescope before 22:00 on Jan. 16 (Saturday). Each question will be given 1, 0.5 or 0 points as follows. If the question is more or less correct it gets 1 point. If it is partly correct it gets 0.5, and if it is missing or completely wrong it gets 0 points.

Make sure to mark the location of each exercise in Gradescope. If this is not done the corresponding exercise(s) will not be graded

1) Consider the method of steepst descent for minimizing the function

$$f(\theta_1, \theta_2) = \theta_1^2 - 2\theta_1\theta_2 + 2\theta_2^2.$$

Derive the gradient of *f*. Show two iterations of the method when it is applied the above function. Assume that α=0.25 and and the starting point is $\theta^{(0)}$ = (1,1). [Provide the values of $\theta^{(k)}$ for k=1,2]

Answer:

What we know:

Minimize the function $f(\theta_1, \theta_2) = \theta_1^2 - 2\theta_1\theta_2 + 2\theta_2^2$ using steepest descent.

The gradient of f is $\nabla f(\theta) = (2\theta_1 - 2\theta_2, -2\theta_1 + 4\theta_2)$

therefore,

alpha = 0.25

$\theta^{(k=0)}$ = (1,1),

Let's find $\theta^{(k=1)}$ and $\theta^{(k=2)}$ with the formula, $\theta^{(k=t)} = \theta^{(k=t-1)}$ – alpha*dt , where dt[0] is the left side of $\theta^{(k=t-1)}$ and dt[1] is the right side.

$\theta^{(k=1)}$:

dt[0] = 2*1 – 2*1 = 0

dt[1] = -2*1 + 4*1 = 2

t[0] = 1 – 0.25 * 0 = 1

Alexander Guðmundsson
alg35@hi.is

t[1] = 1 – 0.25*2 = 0.5

$\theta^{(k=2)}$:

dt[0] = 2*1 – 2*0.5 = 1

dt[1] = -2*1 + 4*0.5 = 0

t[0] = 1 – 0.25 * 1 = 0.75

t[1] = 0.5 – 0.25*0 = 0.5

2) Consider the method of steepest descent applied to some cost function $J(\theta)$ [e.g. the squared error]

a) You observe that the function values decrease very slowly. What can be done to fix that?

Answer:

That usually means the alpha value is too small, to fix it we make the alpha value larger.

b) You observe that the function values *increase* during the first iterations. What can be done to fix that?

Answer:

That usually means the alpha value is too large, to fix it we make the alpha value smaller.

Alexander Guðmundsson
alg35@hi.is

3)

a) When the value of the λ parameter in ridge regression becomes large, the model is at risk of underfitting the data. True? False? Why?

Answer:

TRUE, When the lambda value is too large, the model will be be simple, but you will run the risk of underfitting your data, better said, your model won't learn enough about the training data to make useful predictions.

b) To perform feature selection (identify the most important input variables) while performing linear regression we should use L2 regularization instead of L1 and use a large value of λ instead of a small value. True? False? Why?

Answer:

TRUE, you should use L2 regularization instead of L1 since a predicted feature will get a non-zero coefficient, while L1 can get zero coefficient.

FALSE, like with ridge regression, large lambda gives the risk of underfitting the data.