Week 4 Exercises

The questions below are due on Sunday October 01, 2017; 11:00:00 PM.

You are not logged in.

If you are a current student, please Log In (https://introml.mit.edu/fall17/exercises/ex04?loginaction=login) for full access to this page.

- Videos
  - Week 4, Lecture 1 (https:///introml.mit.edu/lecture_videos/lec_week4.mp4)
- Class Notes for Week 4 (https://introml.mit.edu/__STATIC__/fall17/exercises/ex04/Wk4_notes.pdf)
- Required Exercises

1) INTRO TO LINEAR REGRESSION

So far, we have been looking classification, where predictors are of the form

$$y = sign(\theta^T x + \theta_0)$$

making binary classification as to whether example x belongs to the positive or negative class of examples.

In many problems, we want to predict a real value, such as the actual gas mileage of a car, or the concentration of some chemical. Luckily, we can use most of a mechanism we have already spent building up, and make predictors of the form:

$$y = \theta^T x + \theta_0$$

This is called a *linear regression* model.

We would like to learn a linear regression model from example. Assume X is a d by n array (as before) but that Y is a 1 by n array of floating-point numbers (rather than +1 or -1). Given data (X, Y) we need to find θ, θ_0 that does a good job of making predictions on new data drawn from the same source.

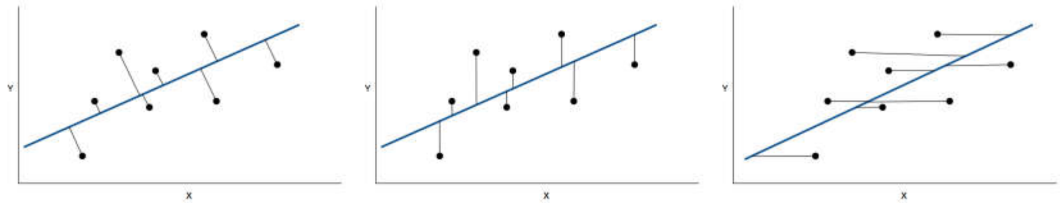
We will approach this problem by formulating an objective function. There are many possible reasonable objective functions that implicitly make slightly different assumptions about the data, but they all typically have the form:

$$J(\theta, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} L(x^{(i)}, y^{(i)}, \theta, \theta_0) + \lambda R(\theta, \theta_0)$$

For regression, we most frequently use *squared loss*, in which

$$L_s(x, y, \theta, \theta_0) = (y - \theta^T x - \theta_0)^2$$

1) Which of the following pictures illustrates the squared loss metric? Assume that the blue line is described by θ, θ_0, the black dots are the (x, y) data, and the light lines indicate the errors.

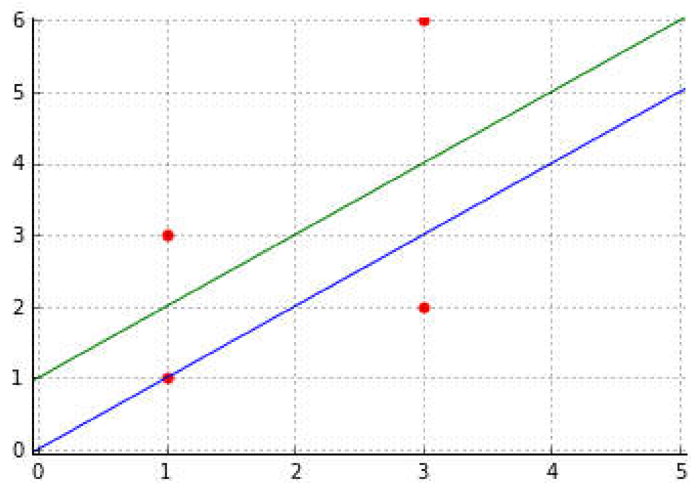


A	B	C

A

2) REGRESSION EXERCISES

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



- The equation of the blue line is: $x - y = 0$
- The equation of the green line is: $x - y + 1 = 0$
- The data points (in x, y pairs) are:$((1,3),(1,1),(3,2),(3,6))$

1) 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]
```

The gradient of the mean squared error regression criterion has the form of a sum over contributions from individual points. The formula for the gradient of the squared error with respect to parameters of a line, θ, θ_0 for a single point (x, y) (without regularizer), is:

$$\left(-2(y - \theta^T x - \theta_0)x, \quad -2(y - \theta^T x - \theta_0)\right)$$

2) What is the gradient contribution from each point to the parameters of the blue line?

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

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

3) 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]
```

4) What is the gradient contribution from each point 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)]
```

5) Mark all of the following that are true:

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

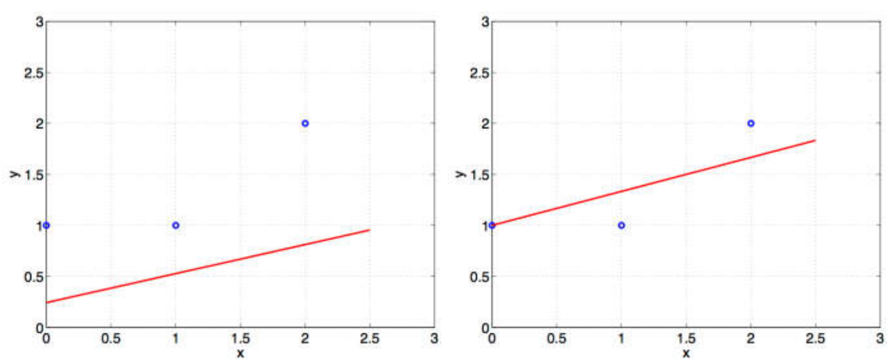
3) RIDGE REGRESSION

If we add a squared-norm regularizer to the empirical risk, we get the so-called *ridge regression* objective:

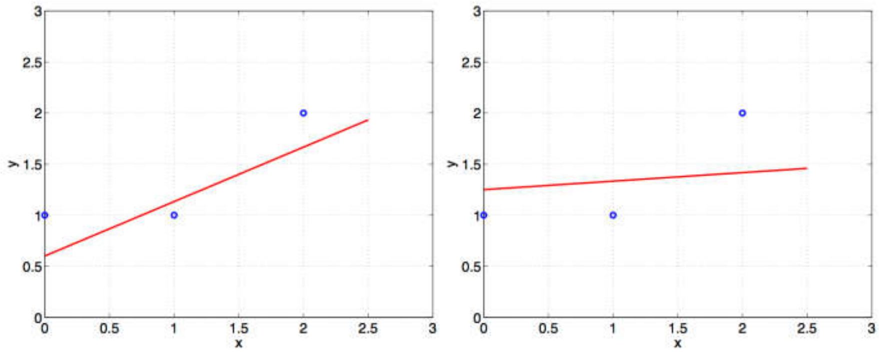
$$J_{ridge}(\theta, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} L_s(x^{(i)}, y^{(i)}, \theta, \theta_0) + \lambda|\theta|^2$$

It's a bit tricky to solve this analytically, because you can see that the penalty is on θ but not on θ_0.

The figures below plot linear regression results on the basis of only three data points $(x^1, y^1), (x^2, y^2), (x^3, y^3)$. 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.



A B



C D

1. $\sum_{i=1}^{3}(y^i - wx^i - w_0)^2 + \lambda w^2$ where $\lambda = 1$ B

2. $\sum_{i=1}^{3}(y^i - wx^i - w_0)^2 + \lambda w^2$ where $\lambda = 10$ D

3. $\sum_{i=1}^{3}(y^i - wx^i - w_0)^2 + \lambda(w^2 + w_0^2)$ where $\lambda = 1$ C

4. $\sum_{i=1}^{3}(y^i - wx^i - w_0)^2 + \lambda(w^2 + w_0^2)$ where $\lambda = 10$ A