Homework1 For Machine Learning

Vergil/Zijun Li李子骏

1. Alice writes a program to play Chess. The program sees win/loss at the end and learns based on that. What type of learning does this represent? (1 point)

It is supervised learning.

It labels the win/lose data and predicts the outcome/future, which is exactly what supervised learning does

2. The amount of snowfall in a given day is measured in inches. Alice designs a learning algorithm to predict the snowfall for the month of March. Is this a classification problem or a regression problem? (2 point)

It is regression problem.

The goal is to predict the snowfall amount in inches which are continuous outcomes(regression), but not predict categorical class labels(Classification)

3. In class, normally we considered cost functions which are squares of the differences between y and h(x). Alice has a new idea: she will use cube of the differences instead. That is, the cost function will be $J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^3$. Does this represent a good cost function? (5 point)

No, it isn't a good cost function.

The loss expresses an error, so it must be always non-negative. And the cube of the difference may result in the difference canceling out.

4. Consider the house price prediction problem from the lecture. Consider the following table of training data. Here x is area in hundred square feet and y is the price in hundreds of thousands of dollars.

x	y
1	2
2	3
4	6
5	7

4.1 Compute J(0, 1) for the above data set. (2 points)

$J(0,1) = \frac{1}{2\times 4}\sum_{i=1}^{4}(h_\theta(x^{(i)}) - y^{(i)})^2$

$J(0,1) = \frac{1}{8} \times ((1-2)^2 + (2-3)^2 + (4-6)^2 + (5-7)^2)$

$J(0,1) = \frac{1}{8} \times (1+1+4+4)$

$J(0,1) = \frac{1}{8} \times 10$

$J(0,1) = 1.25$

4.2 Compute J(1, 1) for the above data set. (2 points)

$J(1,1) = \frac{1}{2\times 4}\sum_{i=1}^{4}(h_\theta(x^{(i)}) - y^{(i)})^2$

$J(1,1) = \frac{1}{8} \times ((2-2)^2 + (3-3)^2 + (5-6)^2 + (6-7)^2)$

$J(1,1) = \frac{1}{8} \times (0+0+1+1)$

$J(1,1) = \frac{1}{8} \times 2$

$J(1,1) = 0.25$

4.3 Compute J(1, 1.1) for the above data set. (2 points)

$J(1,1.1) = \frac{1}{2\times 4}\sum_{i=1}^{4}(h_\theta(x^{(i)}) - y^{(i)})^2$

$J(1,1.1) = \frac{1}{8} \times ((2.1-2)^2 + (3.2-3)^2 + (5.4-6)^2 + (6.5-7)^2)$

$J(1,1.1) = \frac{1}{8} \times (0.01+0.04+0.36+0.25)$

$J(1,1.1) = \frac{1}{8} \times 0.66$

$J(1,1.1) = 0.0825$

4.4 Based on the above, which of the three hypotheses would you choose for prediction and why? (3 points)

I will choose the thrid one $\theta_0 = 1, \theta_1 = 1.1$

Because its lost function has the minimum value which means it mostly fits to the given data set among the three hypotheses.