Julian Tsang

Monday, September 12, 2016 7:49 PM

Machine Learning 2016/2017: Assignment 1

Deadline: September 16

1. Suppose that we have historical data of result of soccer matches of teams playing against Ajax. We wantto use this information to learn to predict at a certain moment whether a team will win, lose or draw against Ajax. Our approach will be based on Machine Learning.
   (a) Define the *given* and the *goal* of the *prediction task* and of the *learning task* that best matches our goal. Classify the learning task as *supervised*, *unsupervised*, *reinforcement learning*, and if supervised as *classification* or *regression*.
   (b) What would be the form of training data for the learning task? Give a small training set.

The given of the prediction task is the historical data of soccer match results.
The goal of the prediction task is to accurately predict whether a team will win, lose or draw against Ajax.
The learning task is classification, which is supervised learning.

The training data for the learning task would use x and y values. X values can be the overall winning record (percentage) of the teams that have played against Ajax. Y values can be based on a value system that classifies winning, losing, or tying (0, 1, 2).

2. **GRADED; 3 points** Given the following data:

X	3	5	6
Y	6	7	10

   (a) Manually (using only a calculator) calculate two iterations of the gradient descent algorithm forunivariate linear regression function. Initialize the parameters such that the regression function passes through the origin (0, 0) and has an angle of 45 degrees. Use a learning rate of 0.1. Give the intermediate results of your calculations and also compute the mean-squared error of the function after 2 iterations.
   (b) Convert the data to z-scores (with mean = 0, sd = 1) and repeat the calculations above. Comparethe results with those for the original data.

$$\theta_0 = 0 \qquad \theta_1 = 1 \qquad h_\theta(\mathrm{x}) = \theta_0 + \theta_1\mathrm{x} = 0 + 1\mathrm{x}$$

Let $\alpha = 0.1$

$$\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

Iteration 1

$$\theta_0 := 0 - (0.1)\frac{1}{3}\left[\left((0+3)-6\right) + \left((0+5)-7\right) + \left((0+6)-10\right)\right] = 0.3$$

$$\theta_1 := 1 - (0.1)\frac{1}{3}\left[\left((0+3)-6\right)(3) + \left((0+5)-7\right)(5) + \left((0+6)-10\right)(6)\right] = 2.4333$$

Iteration 2

$$\theta_0 = 0.3 \qquad \theta_1 = 2.4333 \qquad h_\theta(\mathrm{x}) = \theta_0 + \theta_1\mathrm{x} = 0.3 + 2.4333\mathrm{x}$$

$$\theta_0 := .3 - (0.1)\frac{1}{3}\left[\left((0.3 + 2.4333*3) - 6\right) + \left((0.3 + 2.4333*5) - 7\right) + \left((0.3 + 2.4333*6) - 10\right)\right] = -0.09887$$

$$\theta_1 := 2.4333 - (0.1)\frac{1}{3}\left[\left((0.3 + 2.4333 * 3) - 6\right)(3) + \left((0.3 + 2.4333 * 5) - 7\right)(5) + \left((0.3 + 2.4333 * 6) - 10\right)(6)\right] = 0.3822667$$

Mean Squared Error

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$\theta_0 = -0.09887$ $\theta_1 = 0.3822667$ $h_\theta(\mathrm{x}) = \theta_0 + \theta_1\mathrm{x} = -0.09887 + 0.3822667\mathrm{x}$

$$J(\theta_0, \theta_1) = \frac{1}{2(3)}\left[((-0.09887 + 0.3822667 * 3) - 6)^2\right) + (((-0.09887 + 0.3822667 * 5) - 7)^2) + (((-0.09887 + 0.3822667 * 6) - 10))^2\right]$$
$$= 18.725956$$

b) Converting the dataset into z-scores with the givens (mean 0 and standard deviation 1) would result in a dataset identical to the one given.

3. Suppose that $X1$ predicts Y, with some (mean squared) error MSE. We now extend the data with an additional variable $X2$ and use a learning algorithm that uses both $X1$ and $X2$ to predict Y. What will be the effect on the mean squared error of Y compared to just using $X1$ if $X2$ is equal to: (a) $a + b.X1$
(b) $a + b.X1^2$

$$\frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \qquad h_\theta(\mathrm{x}) = \theta_0 + \theta_1\mathrm{x}$$

a) If $x_2 = a + bx_1$

$h_\theta(\mathrm{x}) = \theta_0 + \theta_1 x_1 + \mathrm{a} + \mathrm{b}x_1 = (\theta_0 + \mathrm{a}) + (\theta_1 + \mathrm{b})x_1$
The hypothesis takes the same form as the original hypothesis function so there should be no effect on MSE.

b) If $x_2 = a + bx_1^2$

$h_\theta(\mathrm{x}) = \theta_0 + \theta_1 x_1 + \mathrm{a} + \mathrm{b}x_1^2 = (\theta_0 + \mathrm{a}) + (\theta_1 x_1 + \mathrm{b}x_1^2)$
The hypothesis becomes a polynomial function so depending on the value of b, it would affect the outcome of MSE. If b = 0, then the hypothesis would be the same form as part (a). Otherwise, MSE would be a smaller value since the polynomial function takes on a more accurate representation of MSE.

4. **GRADED; 2 points** Derive an equation that can be used to find the optimal value of the parameter θ_1 for univariate linear regression without doing gradient descent. This can be done by setting the value of the derivative equal to 0. You may assume that the value of θ_0 is fixed.

$$f = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 = \frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x_1^{(i)} - y^{(i)})^2$$

$$\frac{\partial f}{\partial \theta_1} = \frac{1}{m}\sum_{i=1}^{m}\left(\theta_0 + \theta_1 x_1^{(i)} - y^{(i)}\right)x_1^{(i)} = 0 \ \ (set\ equal\ to\ 0)$$

$$\sum_{i=1}^{m}\theta_0 x_1^{(i)} + \sum_{i=1}^{m}\theta_1 x_1^{(i)^2} + \sum_{i=1}^{m}x_1^{(i)}y^{(i)} = 0$$

$$\theta_1 = \frac{\sum_{i=1}^{m} x_1^{(i)} y^{(i)} - \sum_{i=1}^{m} \theta_0 x_1^{(i)}}{\sum_{i=1}^{m} x_1^{(i)^2}}$$