UNIVERSIDAD POLITÉCNICA DE YUCATÁN

Polynomial Curve Fitting

Classroom Activity 1

Objective

The student will explain the basic concepts of supervised learning, design models according to the special conditions of different types of data and develop solutions for the most common problems presented during the construction of solutions based on Supervised Learning.

Instructions

1. Generate a training set of N = 10 observations of $x = (x_1, \dots, x_N)^T$ with corresponding observations of the values $t = (t_1, \dots, t_N)^T$. Simulate the data using the function $\sin(2\pi x)$ with random noise.
2. Fit the data using a polynomial function of the form:

$$y(x, w) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

where M is the order of the polynomial, and x^j denotes x raised to the power of *j*. The polynomial cofficients $w_0, w_1, \dots, w_M$ are collectively denoted by the vector *w*. The values of the coefficients can be determined by fitting the polynomial to the training data. This can be done by minimizing an error function that measures the misfit between the function *y(x, w)*, for any given value of *w*, and the training set data points. Use the following Error Function:

$$E(w) = \frac{1}{2}\sum_{n=1}^{M}\{y(x_n, w) - t_n\}^2$$

   a. Proof that the w can be obtained using the following:

$$w^* = A^{-1}T$$

   Please, look your class notes for the definition of A and T.

   b. Create a function to fit your data using a Polynomial Function. Name your function as get_w(x, t, M). The inputs must be *x*, *t* and *M*. The output must be the *w* vector.

3. Fit your data with different values of M (M = 0, 1, 3, 6, 9). Make a table of the order M and its corresponding Root-Mean-Square Error. The Root Mean Square Error can be obtained with:

$$E_{RMS} = \sqrt{2E(w^*)/N}$$

w* is the w vector that minimizes E(w).

4. Make a new simulation with N = 20. Select randomly 70% of the data to be used as training set and use the other 30% of the data as the evaluation set.
   Make a comparison (plot) of the E_{RMS} in the training and test set with different values of M (Fig. 1.5, Bishop et.al. 2006).
5. Make two simulation with $N_1 = 15$ and $N_2 = 100$. Fit a Polynomial Function of M = 9. Make a graphical comparison and include a table with its corresponding E_{RMS}(Fig. 1.6, Bishop et.al. 2006).
6. The dataset auto.csv includes the characteristics of different cars (mpg, cylinders, displacement, horsepower weight, acceleration, year origin, name). Fit a polynomial curve between:
   - Mileage as function of the Horsepower.
   - Mileage as a function of the Displacement.
   - Horsepower as function of the Weight.

   For each pair:
   a. Make a scatterplot of the variables.
   b. Select randomly 70% as the Training Set and the other 30% as the Test Set. Fit a Polynomial Function. Select an appropriate value of M. Don't forget to include the E_{RMS}.

Homework Submission

Jupyter Notebook File
Submit your Jupyter Notebook.
Deadline: January 23rd, 2019 at 23:59 hrs.

Grading Scheme

Check list

Task 1	14 pts
Task 2	14 pts
Task 3	14 pts
Task 4	14 pts
Task 5	14 pts
Task 6	30 pts