MALAVIYA NATIONAL INSTITUTE OF TECHNOLOGY, JAIPUR

Department of Computer Science and Engineering

M.Tech Programming Lab 2019

Exercise Sheet

Note:

(1) All the problems are to be implemented in Python. The exercises are to be done individually.

(2) The submission date is the last lab day for the exercise. The number of labs for each exercise is mentioned alongside.

1. (10 points, Labs - 1) Write a python script that can find w_0 and w_1 for an arbitrary dataset of number of hours studied versus rank of a students as $\{(x_n, y_n)\}$ pairs. Find the linear model, $y = w^T x$, that minimizes the squared loss. Derive the optimal parameter value, w, for the total training loss: $L = \sum_{n=1}^{N}(y_n - w^T x_n)^2$. Using the model predict the rank for the number of hours studied.
   Load the data stored in the file syntheticdata.mat.
   Fit a 4^{th} order polynomial function $f(x; w) = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4$ to this data. What do you notice about w_2 and w_4?
   Fit a function $f(x;w) = w_0 + w_1x + w_2 \sin(\frac{(x-a)}{b})$ to this data, assuming a and b are fixed in some sensible range. Show a least square fit using this model. What do you notice about w_1 and w_2. Comment about generalization and overfitting.