German International University of Applied Sciences
Informatics and Computer Science
Dr. Caroline Sabty
Eng. Nadeen Amr

Machine Learning, Winter 2022
Mini Project 1
Due date is November 15, 2022 at 11:59 PM
Submitted in groups of maximum 2

Using Polynomial Regression allows our model to go from the equation in the form (shown here as if we only had one x feature):

$$\hat{y} = \beta_0 + \beta_1 x_1 + \epsilon$$

and create more features from the original x feature for some **d** degree of polynomial.

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_1 x_1^2 + ... + \beta_d x_1^d + \epsilon$$

Obviously we need to be careful about choosing the correct value of **d**, the degree of the model.

When having multiple X features, not just a single one as in the formula above, the PolynomialFeatures model also takes interaction terms into account for example, if an input sample is two dimensional and of the form [a, b], the degree-2 polynomial features are $[1, a, b, a^2, ab, b^2]$.

For this assignment, using the Advertising dataset, you are asked to produce a polynomial regression model that best suits the given dataset by trying different higher order models (maximum x^{10}). Upon choosing the optimal model, compare its performance measure outputs to those of simple linear regression and ridge regression with cross validation.

A few things to take into consideration as you are working:

a) Use the train-test-split method in such a way that each model is given the same portion of the data for training and testing each time

b) Produce an error-complexity plot for all the models and base your choice of polynomial regression optimal model on its output, clarifying the reasoning in the comments

c) Upon determining the optimal model, you need to save it into a file and call it "optimal_polyreg_model"

d) The final comparison for the simple linear regression, ridge regression with cross validation and polynomial regression should be done using the evaluation metrics discussed in tutorial 3 as well as the predicted sales for the campaign with the following expenditure: 149k on TV, 22k on Radio, and 12k on Newspaper Ads

Deliverables:

a) Your code is to be submitted to ics504ml@gmail.com as a zip file containing your notebook (include the names and IDs of the team members in the body of the email with subject "Mini Project 1").

PLAGIARISM IS NOT TOLERATED AND COPIED WORK WILL BE AWARDED 0 POINTS FOR BOTH TEAMS INVOLVED! **There will be an individual evaluation for each team**