Deep Learning and Computer Vision
Overfitting, Regularization and Cross-validation

In this introduction exercise, we are going to work on a polynomial ridge regression task. The idea is to understand the notions of overfitting, regularization and cross-validation.

We have some points in a 2D space (X,y), which are our training data. We want to learn a polynomial function P() that predicts the y from the X : y = P(X).

The test data will beprovided when you think you have a « good » predictor.

The optimization step can not be changed, you can only play with two hyper-parameters: the order of the polynomial function and the regularization term.

You can use your own python environment or just run your code online on trinket (https://trinket.io/python3).

1/ Read, understand and run the provided code.

2/ What is the R2 score (maximum, zero value) ?

3/ For the orders from 1 to 8, which polynomial function provides the best training results (regularization coefficient = 0, here) ? Why ?

4/ By fixing the order to 5, play with the regularization coefficient from 1e-4 to 10. What is the impact on the training result ? Why ?

5/ Use the « cross_val_predict » function to run a cross validation. What would be a good number of folds, here ?

6/ Test a cross validation prediction with an order of 5 and a regularization of 0.1. Observe the result. How many values contains this vector ?

7/ Use the cross-validation to find the best hyperparameters.

8/ Test your solution on the the test data.