

Assignment 1. Learning model of minimal capacity

Given a set of **inputs** x and corresponding **outputs** y of an object, a learning model of minimal capacity should be identified.

An assumption is made that the function to be learned has a minimal number of degrees of freedoms. So, the task is to find the regression model which is not overfitted or underfitted.

This can be done by using the Principal Component Analysis to find out the most significant features of the learning model based on the given dataset.

1. Given is the following dataset (the training data):

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

2. Plot the graph $y(x)$.
3. Find the mean values of the both x , y .
4. Calculate the covariance (2x2) matrix.
5. Find the eigenvalues and eigenvectors of the covariance matrix.
6. Compare the vectors to see if there is a vector that can be identified as the principal component.
7. Create a learning (regression) model utilizing the principal component.
8. Plot the graph $y = f(x)$ representing this new model.
9. Use the model to test it for the unused, so far, data. What output the trained model will suggest for $x = 2.5$ and 5 ?

Show all the steps. You may use MATLAB, Python and other packages to help you with calculations.

Provide the analysis and explanation of the techniques used.

Prepare the pdf file with your results.

You can work in small groups up to three students.

Submit to the Beachboard by September 20, 11:59pm.

Also, bring 2 (two) printed copies of your report to be handed to a reviewing team.

The assigned reviewing team will present brief comments on the submitted report to the next class.