

# MACHINE LEARNING FOR ROBOTICS I - 86928

[Dashboard](#) / [My courses](#) / [86928](#) / [Lab assignments](#) / [Lab 2: Linear regression](#)

## Lab 2: Linear regression

- Task 1: Get data
- Task 2: Fit a linear regression model
- Task 3: Test regression model

Describe everything in a report. Then submit everything.

### Task 1: Get data

Download the [turkish-se-SP500vsMSCI](#) and the [mtcarsdata-4features](#) data sets.

Do what is necessary to make them readable in Matlab, for instance with the `load` function or the `csvread` function

Original sources of the data:

- The [Turkish stock exchange data](#) can be downloaded from the [U.C.I. Machine Learning Repository](#) along with many others.
- The MT cars data are available as the command "mtcars" in the (open source) [R statistical/data analysis language and environment](#), along with many other example data sets [Here is the documentation](#).

### Task 2: Fit a linear regression model

Using the slides, reproduce the examples that you have seen during the lectures:

1. One-dimensional problem without intercept on the Turkish stock exchange data
2. Compare graphically the solution obtained on different random subsets (10%) of the whole data set
3. One-dimensional problem with intercept on the Motor Trends car data, using columns mpg and weight
4. Multi-dimensional problem on the complete MTCars data, using all four columns (predict mpg with the other three columns)

**Remark:** A random set of  $m$  unique indices from 1 to  $N$  is obtained in Matlab as follows:

```
allIndices = randperm(N);  
randomSubset = allIndices(1:m)
```

or in one line:

```
randomSubset = randperm(N,m)
```

In python with `import numpy as np` it would be just one line:

```
randomSubset = np.random.permutation(N)[:m]
```

**Remark:** To make differences more visible, rather than random sets you can select observations from different ends of the data set, i.e. from the beginning and from the end.

The idea is that, since the data are collected across time (a multidimensional time series), data collected in similar periods may be more similar than data collected from the beginning and the end of the whole period, which is one year.

As for random data, these will sample the whole set and will likely contain some instances from the whole period. What we are capturing with regression is the average behaviour of a data set. It is very likely that the average behaviour of two random data subsets will be similar to the average behaviour of the whole set; you may try it to convince yourself experimentally.

### Task 3: Test regression model

---

Re-run 1,3 and 4 from task 2 using only 5% of the data.


Compute the objective (mean square error) on the training data

Compute the objective of the same models on the remaining 95% of the data.

Repeat for different training-test random splits, for instance 10 times. Suggestion: write code for this job, don't waste time repeating manually. Matlab scripts are done for that.

Show the results (using a graph or a table of values) and comment.

### Submission status

<b>Attempt number</b>	This is attempt 1.	
<b>Submission status</b>	Submitted for grading	
<b>Grading status</b>	Not graded	
<b>Last modified</b>	Thursday, 7 July 2022, 2:16 PM	
<b>File submissions</b>	 <a href="#">Machine Learning Lab2_S4845876.zip</a> 7 July 2022, 2:16 PM	
<b>Submission comments</b>	<a href="#">▶ Comments (0)</a>	

[Edit submission](#)[Remove submission](#)

You can still make changes to your submission.

[◀ Lab 1: Naive Bayes classifier](#)[Lab 3: kNN classifier ▶](#)[Condizioni d'uso](#)[Policy](#)[Siti AulaWeb](#)[Help](#)