

Lab 5: Regression and Dimensionality Reduction (2%)

Lab 5 builds on top of Lab 4 by introducing **dimensionality reduction** into the process of training and comparing predictive models (i.e., classification and regression models). In general, the more features/columns a dataset has (with a fixed number of examples), the more difficult might be to train an accurate predictive model. For datasets with a large number of features, it is almost always necessary to either

- select the most important features or
- transform the features to reduce the dimensionality of the feature space

before training a model. Similar to data preparation dimensionality reduction needs to be done separately for the training and the test sets.

The first goal of Lab 5 is to understand how to introduce dimensionality reduction techniques into the training pipeline. This time we will use regression algorithms (i.e., algorithms for numeric prediction) instead of classifiers, but the same dimensionality reduction techniques can be used in classification pipelines. **The second goal is to practice training regression models.**

Task 1

- Download the following files:
 - [Lab 5 - Regression and Dimensionality Reduction.ipynb](#) - This is the example notebook to use in the lab.
 - [fish.csv](#) - This is the dataset used in the example notebook.
 - [insurance.csv](#) - This is the dataset to use in Tasks 2 and 3.
-

Task 2

- Repeat the experiment in *Lab5 - Regression and Dimensionality Reduction.ipynb* with *insurance.csv* and draw conclusions from the results you observe in a markdown cell. Add a third regression algorithm of your choice. Describe how it compares to Random Forest and Linear regression in a markdown cell (**max 250 words in a markdown cell**).
-

Task 3

- Add two dimensionality-reduction methods which are not used in *CS5062 - Lab5 - Regression and Dimensionality Reduction.ipynb* to the training pipelines

for *insurance.csv*. Aim at dimensionality reduction techniques that are sufficiently different from the ones used in the example notebook. Draw conclusions from the results you observe in a markdown cell (**max 250 words in a markdown cell**).

Submission

Save your Jupyter notebook and name it *lab5ipynb*.

Submit your notebook in the Course Tools > Assignment section by 22-Mar-2024 23:59. Late submissions will not be accepted.