

INDIVIDUAL WORK: Deadline: 28/04/2021

Sent the pdf file with the work AND source codes by email to
leandro.coelho@pucpr.br **SUBJECT:** Computational work 01

Computational work 01:

Regularization approaches

Leandro dos Santos Coelho

Pontifícia Universidade Católica do Paraná (PUCPR), Escola Politécnica
Pós-Graduação em Engenharia de Produção e Sistemas (PPGEPS), Graduação em Engenharia de Controle e Automação (Mecatrônica)
Rua Imaculada Conceição, 1155, CEP 80215-901, Curitiba, PR, Brasil

Universidade Federal do Paraná (UFPR), Graduação e Pós-Graduação em Engenharia Elétrica, Campus Centro Politécnico
Av. Cel. Francisco H. dos Santos, 100, CEP 81530-000, Curitiba, PR, Brasil

e-mail: leandro.coelho@pucpr.br; lscoelho2009@gmail.com

Currículo Lattes: <http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4792095Y4>

Google Scholar: <https://scholar.google.com/citations?user=0X7VkC4AAAAJ&hl=pt-PT>

Linkedin: <https://www.linkedin.com/in/leandro-dos-santos-coelho-07a08893/>

Computational work

INDIVIDUAL WORK: Deadline: 28/04/2021

Sent the pdf file with the work AND source codes by email to leandro.coelho@pucpr.br **SUBJECT:** Computational work 01






- 1) **Choice three** case studies with focus in **regression**.
- 2) **Adapt** the **linear regression** (with OLS), **ridge**, **LASSO** and **elastic net** approaches to the **three** case studies and after run it (in Matlab, R or Python).
- 3) **Analysis** (tabels, figures, statistical analysis, box-plot, histograms, train-validation performance, etc) of the results for forecasting (MSE or MAPE or RMSE or R^2 metrics using **cross-validation**, **5-fold**) to **diferent values** of regularization parameter and **without/with a feature selection method**.

Dataset (adopt case study with 1 output)



98 Data Sets

Table View [List View](#)

Name	Data Types	Default Task	Attribute Types	# Instances	# Attributes	Year
 3D Road Network (North Jutland, Denmark)	Sequential, Text	Regression, Clustering	Real	434874	4	2013
 Air Quality	Multivariate, Time-Series	Regression	Real	9358	15	2016
 Air quality	Multivariate, Time-Series	Regression	Real	9358	15	2016
 Airfoil Self-Noise	Multivariate	Regression	Real	1503	6	2014
 Amazon Access Samples	Time-Series, Domain-Theory	Regression, Clustering, Causal-Discovery		30000	20000	2011

Dataset (adopt case study with 1 output)


Penn Machine Learning Benchmark (PMLB) repository

The screenshot shows the GitHub repository for EpistasisLab / penn-ml-benchmarks. The repository has 13 users, 25 watchers, 432 stars, and 71 forks. The main branch is master. The repository contains a directory named 'regression' which lists 10 datasets, each with a folder name, a description 'add regression', and a commit date '2 years ago'.

Dataset Name	Description	Commit Date
1027_ESL	add regression	2 years ago
1028_SWD	add regression	2 years ago
1029_LEV	add regression	2 years ago
1030_ERA	add regression	2 years ago
1089_USCrime	add regression	2 years ago
1096_FacultySalaries	add regression	2 years ago
1191_BNG_pbc	add regression	2 years ago
1193_BNG_lowbwt	add regression	2 years ago
1196_BNG_pharynx	add regression	2 years ago
1199_BNG_echoMonths	add regression	2 years ago

Dataset (adopt case study with 1 output)

OpenML


 OpenML


Search


HELP


SIGN IN


Explore


 Data


 Task


 Flow


 Run

 Study

 Task type

 Measure

 People

 OpenML Regression 30

Created 29-05-2018 by [Adithya Balaji](#) Visibility: public

SEARCH THESE DATA SETS IN MORE DETAIL


DESCRIPTION

30 DATA SETS


0 TASKS

0 FLOWS


0 RUNS

 **Moneyball (2)**

In the early 2000s, Billy Beane and Paul DePodesta worked for the Oakland Athletics. While there, t...
★ 0 runs ♥ 0 likes ☁ 7 downloads 📶 7 reach ⚡ 3 impact
1232 instances - 15 features - 0 classes - 3600 missing values

 **house_16H (1)**

This database was designed on the basis of data provided by US Census Bureau [http://www.cens...
★ 0 runs ♥ 1 likes ☁ 6 downloads 📶 7 reach ⚡ 5 impact
22784 instances - 17 features - 0 classes - 0 missing values

 **water-treatment (1)**

1. Title: Faults in a urban waste water treatment plant 2. Source Information: – Creators: Manel Poc...
★ 0 runs ♥ 0 likes ☁ 1 downloads 📶 1 reach ⚡ 5 impact

Quote

Absence of evidence is **not** evidence of absence.

Carl Sagan (1934-1996)

American astronomer, cosmologist, astrophysicist, astrobiologist.

Sagan published more than 600 scientific papers and articles and was author, co-author or editor of more than 20 books.

Sagan advocated scientific skeptical inquiry and the scientific method, pioneered exobiology and promoted the Search for Extra-Terrestrial Intelligence (SETI).

