# EP3260: Machine Learning Over Networks
# Peer-review of CA1 of group 2

Stefanos Antaris[*1], Amaru Cuba Gyllensten[†1,2],
Martin Isaksson[‡1,2,3], Sarit Khirirat[§1], and Klas Segeljakt[¶1,2]

[1]*KTH Royal Institute of Technology*
[2]*RISE AI*
[3]*Ericsson Research*

February, 2019

## Contents



## 1 Computer assignment CA1

### 1.1 General comments

- Documenting this computer assignment in a Jupyter notebook is a good choice as it makes it easy for the reviewer to read and understand what has been done. Excellent!

- It would have been nice to include some more comments between the cells as well.

[*]antaris@kth.se
[†]amaru.cuba.gyllensten@ri.se
[‡]martisak@kth.se
[§]sarit@kth.se
[¶]klasseg@kth.se

### 1.2 Find a closed-form solution for this problem

- The derivation of the close-form solution is good and correct.

### 1.3 Find the optimal linear regressor from the closed-form expression

- There is no link to the dataset or explanation on how to find this, neither is there an overview of its contents. This means we can't reproduce this example.
- Since we have not seen the dataset, we can't know why you are picking columns with `X = data.iloc[:, 5:127]`. Presumably, these are columns such as `state` and `county`, but should have been made clearer.
- The size of the identity matrix contains a magic number and should depend on the dataset shape.
- Even though it was not part of the assignment, $\lambda$ was set to 0, which was a little bit of a boring choice :).

### 1.4 Repeat for Individual Household Electric Power Consumption dataset

- In this example you have printed the head of the dataset, which makes it a little bit clearer what happens in the beginning.
- You are reusing the same closed form, so it would have made sense to break the calculation out to a new function.

### 1.5 How would you address even bigger datasets?

- Interesting notes here, there might be other ways too, such as gradient descent methods.