

# **Machine Learning: Homework on Regularized Linear Regression and SVM for Classification**

Fabio Vandin

December 5<sup>th</sup>, 2019

# Instructions

You have to complete 2 Jupyter notebooks:

- one for regularized regression (house sale data)
- one for SVM on classification (digits images data - MNIST)

Both notebooks have missing code: need to fill in what is missing

You also need to write some text (to explain choices or describe results)

Feel free to add cells with text if you need to explain or describe some ``non-standard'' decision!

**FIRST THING TO DO: you need to put your ID number in both notebooks (as seed for random number generators).**

**IMPORTANT: code already there is a guideline, if you want to change or rewrite some/all of it go ahead, but make sure to answer all TO DOs!**

# Deadline

Submit your completed notebooks:

- **deadline: Thursday December 19<sup>th</sup>, 11:55 PM**
- **use link in elearning website**

Submit 2 files (1 completed notebook for regularized regression, 1 completed notebook for SVM on classification) - **Only submit your completed notebooks!**

**IMPORTANT:** Use the following file names for the 2 files that you have to submit:

- for the regularized regression notebook:  
RR\_FirstnameLastName\_IDnumber.ipynb
- for the SVM classification notebook: SVM\_FirstnameLastName\_IDnumber.ipynb

**Example:** student Fabio Vandin, ID number 000001 will submit files:

- RR\_FabioVandin\_000001.ipynb
- SVM\_FabioVandin\_000001.ipynb

**WRONG FILE NAMES = 0 POINTS**

In lab assistance: **Thursday December 12<sup>th</sup>, room Te (and Ue if needed),  
4:30-6:15pm**

# Dataset for regularized regression

## Regularized regression problem:

- house sale prices for King County  
<https://www.kaggle.com/harlfoxem/housesalesprediction>

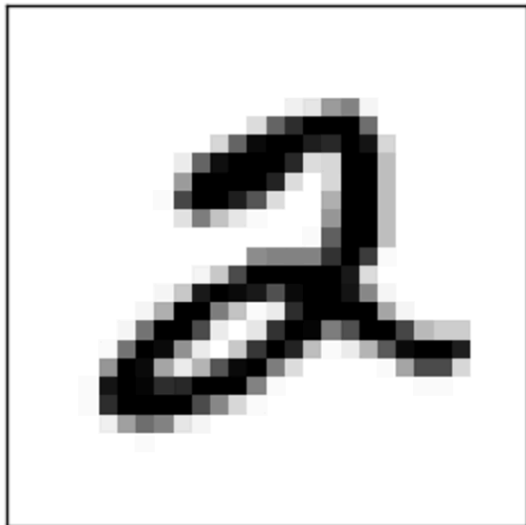
**See link above for descriptions and references  
(links are also in the notebook)**

# Dataset for SVM classification

MNIST dataset:

- <http://yann.lecun.com/exdb/mnist/>
- <http://mldata.org/repository/data/viewslug/mnist-original/>
- 70000 training instances, input =vector of 784 integer values in  $[0, \dots, 255]$  = 28x28 matrix

INPUT:



TARGET/LABEL: 2