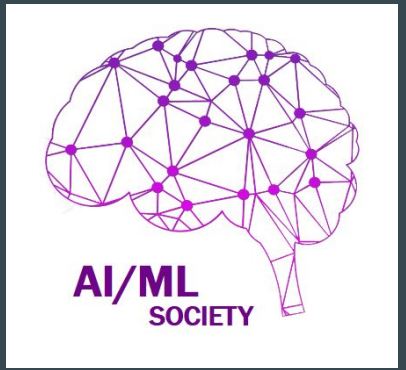
University of Manchester AI ML Society - Introduction to ML

...

Workshop 2
16th of October 2019



By Botond Maros

Intro to ML timetable

Workshop 1 - Introduction to Machine Learning

Workshop 2 - Data preprocessing

Workshop 3 - Fundamental Algorithms I

Workshop 4 - Fundamental Algorithms II

Workshop 5 - Neural Networks Part I

Workshop 6 - Neural Networks Part II

Today's session

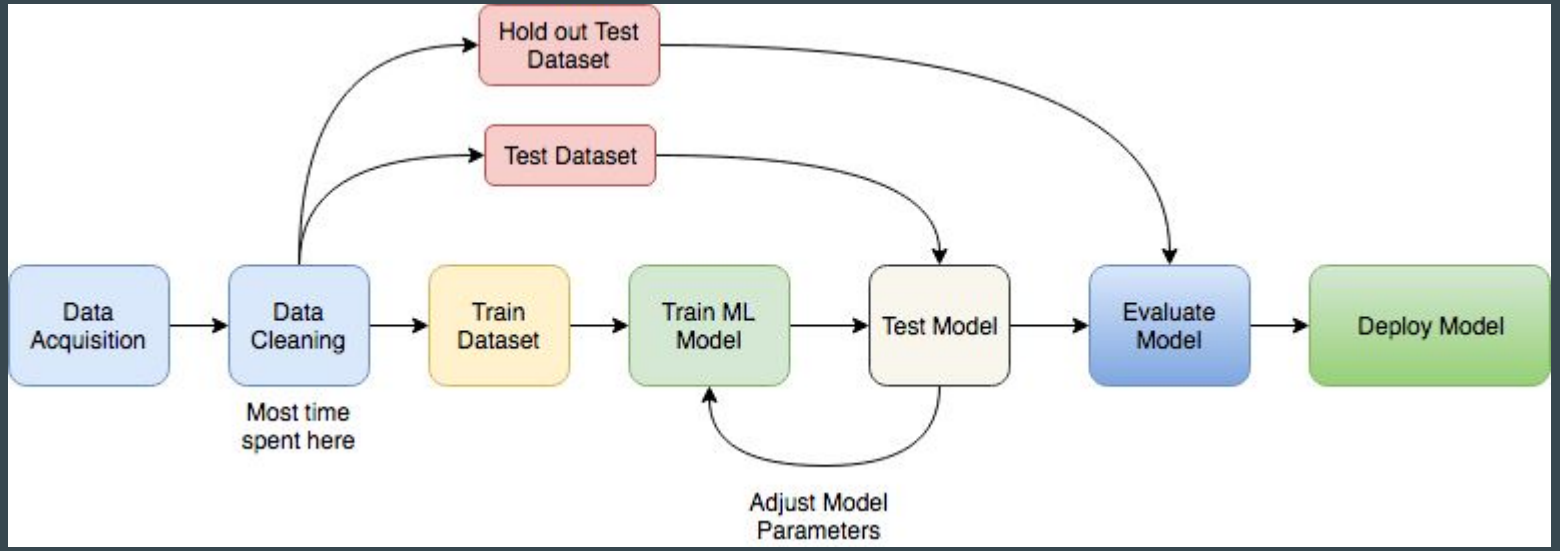
Recap of workshop 1

What is data preprocessing and why is it important?

Coding exercise

Take home challange

Recap of the ML Workflow



http://www.shanklab.com/machine-learning-getting-started/

Machine Learning Workflow in Practice

- Common misconception: what is useful and what is trendy in practical research and industry
- In the “wild” massive, unstructured data sets
- Problem ambiguity -> visualize

What to look out for in numerical data

- Number of samples per class (class imbalance)
- Number of total samples in dataset (choosing models)
- Number of features (dimensionality)
  - Categorical vs numerical features
  - Feature dependency
- Missing features (average, median, etc.)

How to represent non numerical data

- Text
  - create corpus
  - sparse matrix
  - numbers represent occurrence
  - every row is a training example
- Images:
  - colors -> numbers
  - use matrices
  - every matrix is a training example
  - flat matrix to vector

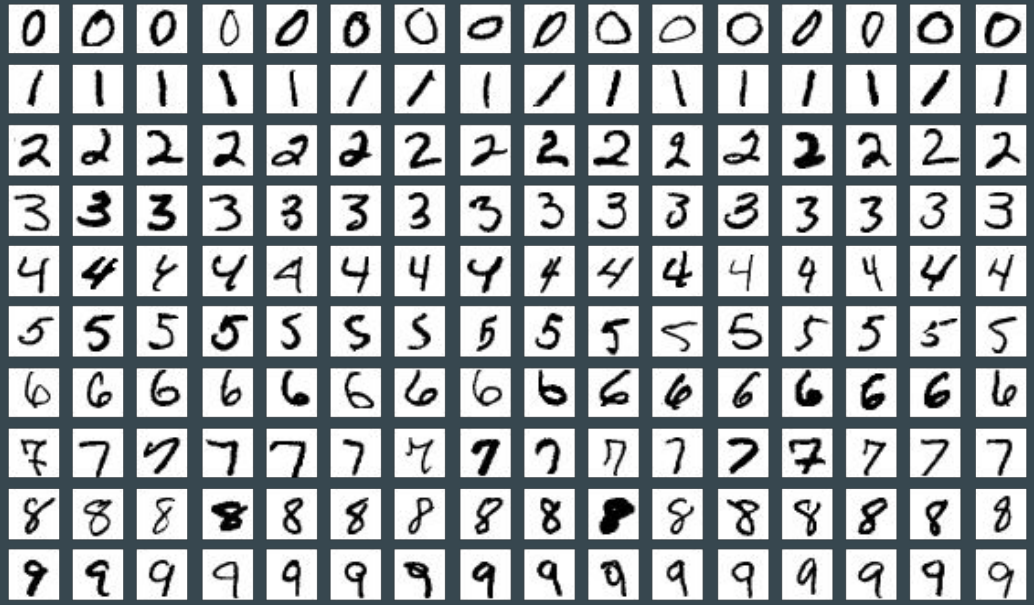
Evaluating your model

		Predicted	
		Negative	**Positive**
Actual	**Negative**	True Negative	False Positive
	Positive	False Negative	True Positive

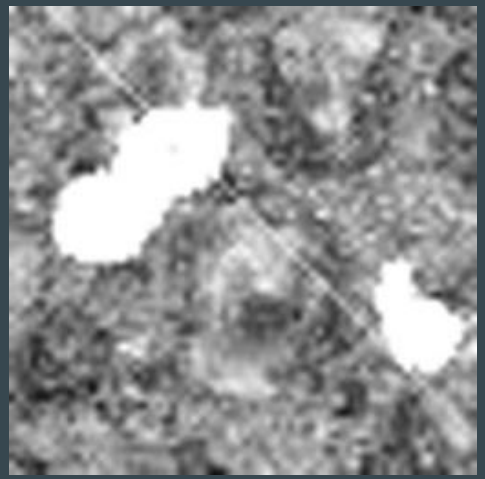
Evaluating your model

- Recall = True positive / (True positive + False negative)
- Precision = True positive / (True positive + False positive)
- F1 = 2 *Precision *Recall / (Precision + Recall)

MNIST example



Modified MNIST



Thank you!