DATAMAX

Introduction to Machine Learning

DATAMAX

ML WORKSHOP

Today's Agenda

1. Introduction
2. Python 101
3. Intro into ML
4. Linear Regression - Theory
5. Linear Regression - Practice

01

Introductions

BUJAR BAKIU

Bujar Bakiu

- MD & Machine Learning Engineer, Data Max
- Graduated 2016, MSc, RWTH Aachen University
- Passionate about Machine Learning, Software Engineering and running
- Find me in Twitter: @bujarbakiu
  LinkedIn: linkedin.com/in/bbakiu
  Email: bujar@data-max.io

02

Python 101

03

Intro into ML

ML WORKSHOP

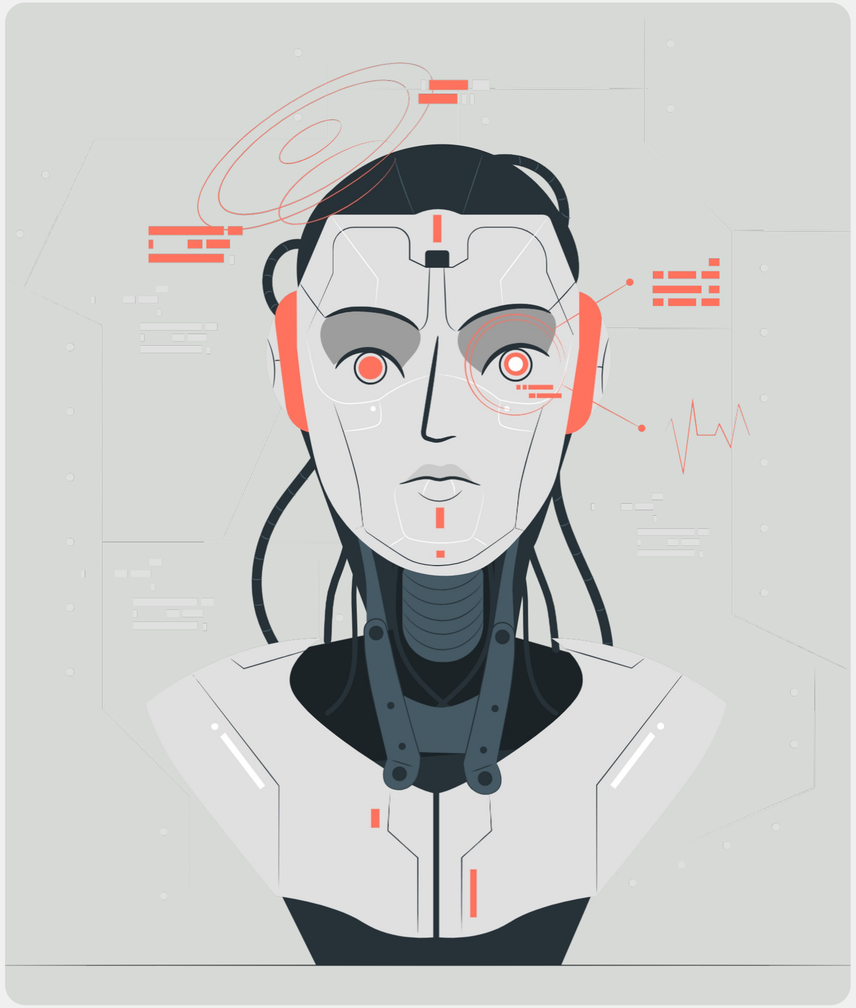
What is Machine Learning

- **Goal**

  Machines that *learn* to *perform* a *task* from *experience*

- **Why?**

  - Task cannot be well defined
  - Volume of data is just too big to analyze
  - Environment changes
  - New knowledge become available

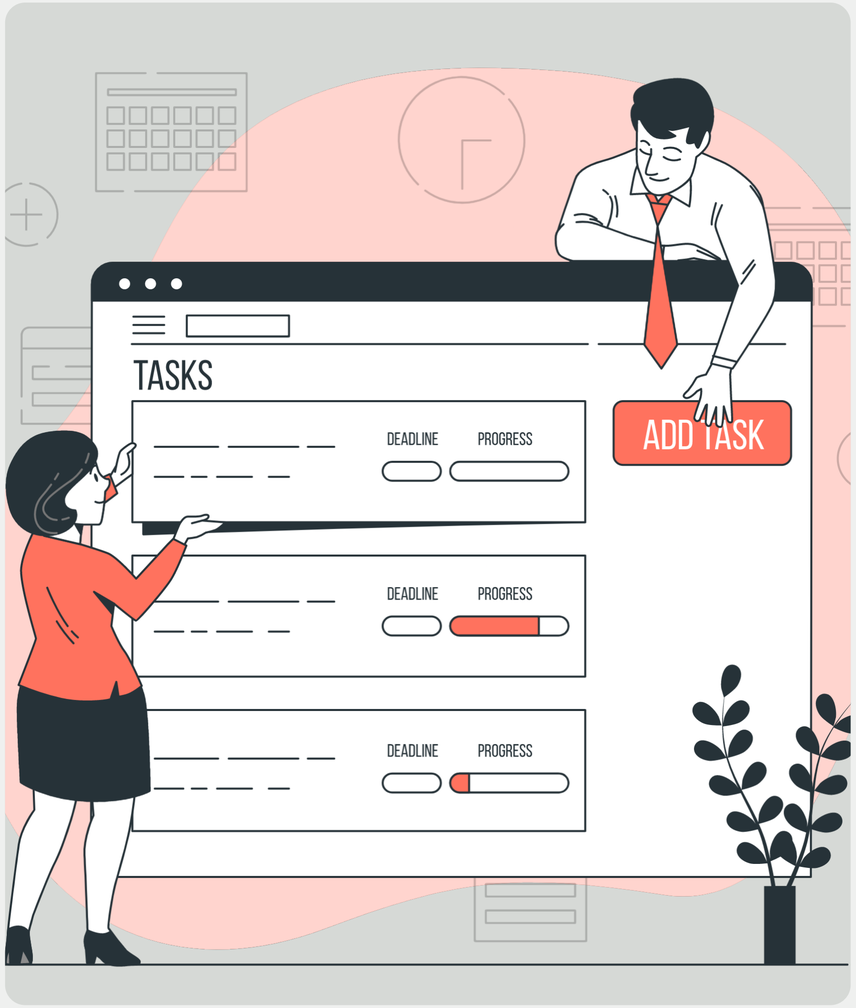


ML WORKSHOP

Learning

Machines that *learn* to *perform* a *task* from *experience*

- *Learning* to perform a task from experience
- We don’t want to encode knowledge
- Machine should discover and learn automatically from past
- Statistics, Probability Theory, Decision Theory, Information Theory, Optimization Theory

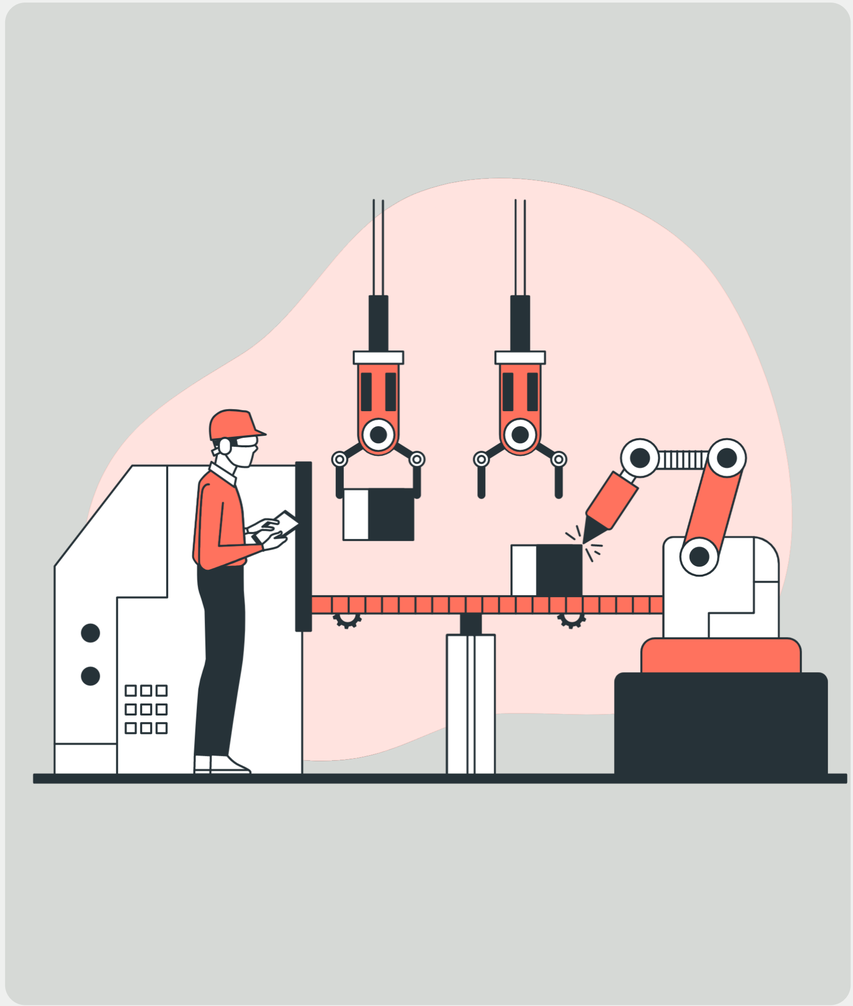


M L W O R K S H O P

Task

Machines that *learn* to *perform* a *task* from *experience*

- Usually, a function
  y = f(x, w)
- X: input aka. features
  W: parameters aka. weight
- y: output aka, target (label)
- Classification vs. Regression vs. Reinforcement Learning

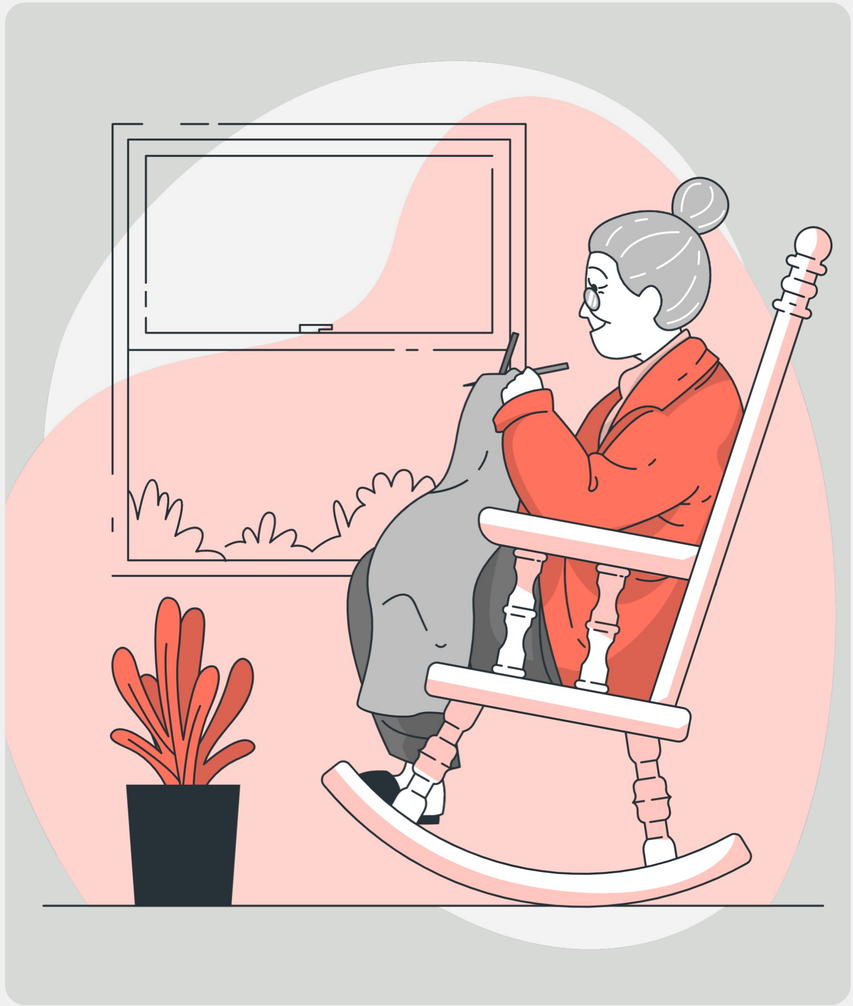


ML WORKSHOP

Perform

Machines that *learn* to *perform* a *task* from *experience*

- “99% correct classification”
  - Of what?
  - What was the training data set? What about testing?
  - Is it matching real world?
- Performance Measurement:
  - Numbers
  - One or more metrics
  - Different for classification vs. regression

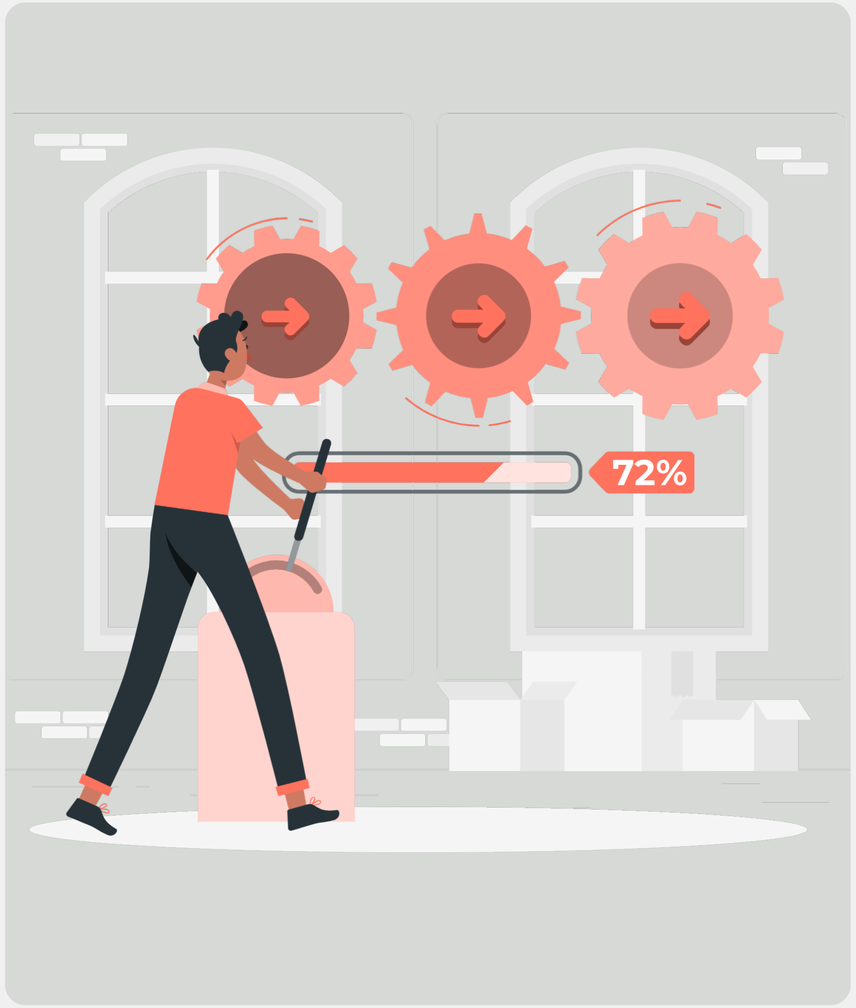


ML WORKSHOP

Experience

Machines that *learn* to *perform* a *task* from *experience*

- When labelled data available:
  - Supervised Learning
- No labelled data:
  - Unsupervised learning
- Some labelled data:
  - Semi-supervised learning
- Feedback/rewards:
  - Reinforcement Learning



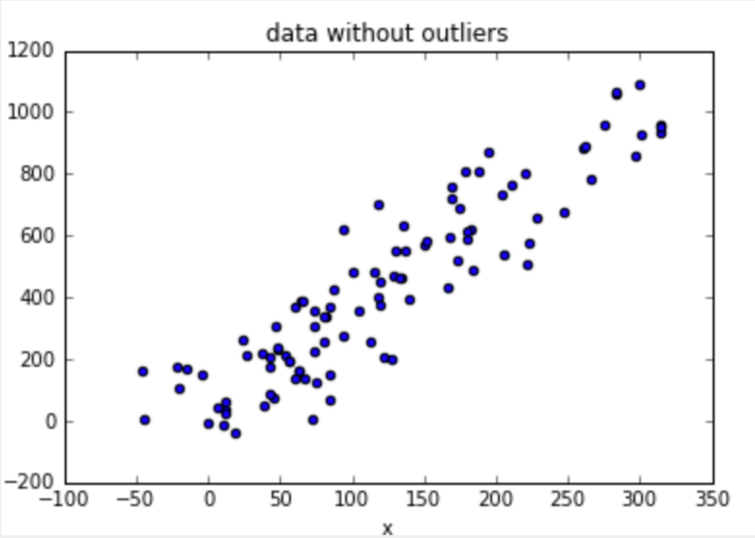
ML WORKSHOP

Basic Algorithms

- Linear Regression
- Logistic Regression
- SVM
- Decision Tree / Random Forest
- Naïve Bayes
- K-Nearest Neighbours
- K-Means Clustering
- DBSCAN Clustering

04

Linear Regression

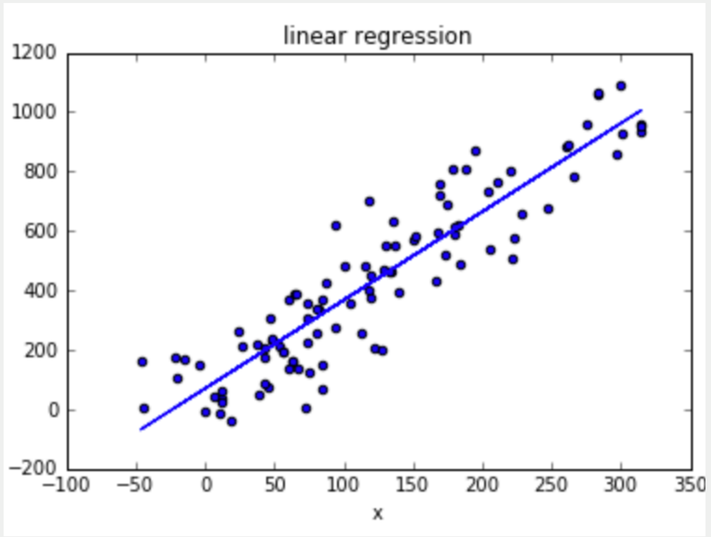


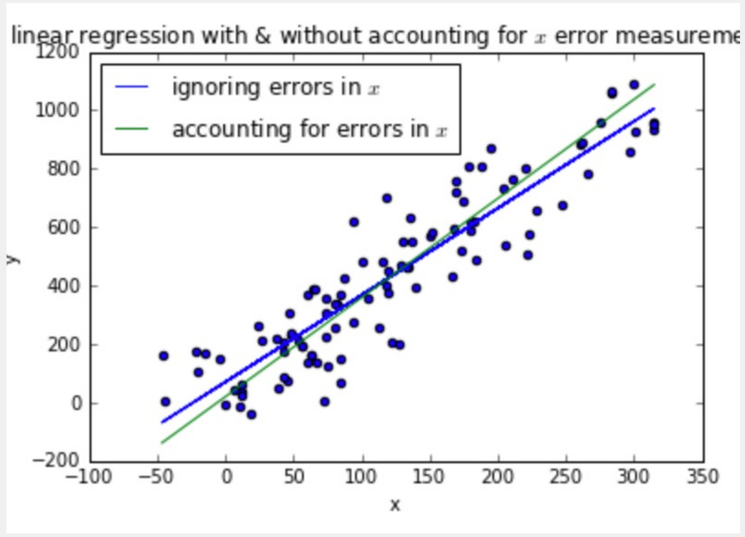
M L W O R K S H O P

Linear Regression

Relationship between target and features

Which line best fits the data?





ML WORKSHOP

Linear Regression

$$y = wx + w_0$$

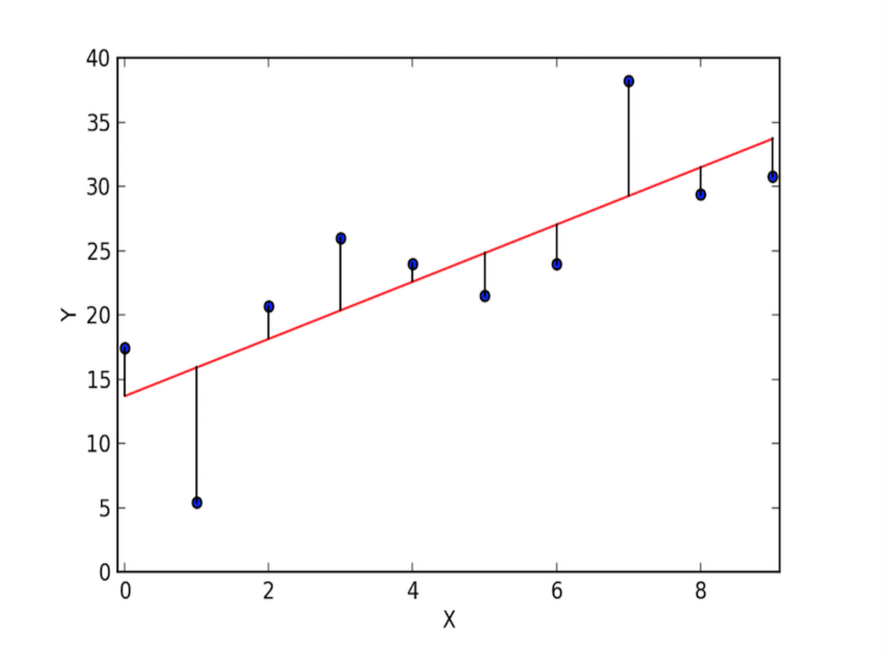
y -> target

x -> feature

w -> weight

$$y = f(x_1, x_2, x_3, x_4, \ldots, x_n)$$

$$y_k = w_1x_{1k} + w_2x_{2k} + w_3x_{3k} + w_4x_{4k} + \ldots + w_nx_{nk} + w_0$$



M L W O R K S H O P

Linear Regression

- Goal:
  Reduce residuals (loss)
- Least square method
  - Try to minimize the sum-of-squares error
- LS

$$E(\mathbf{w}) = \sum_{n=1}^{N} (y(\mathbf{x}_n; \mathbf{w}) - \mathbf{t}_n)^2$$

05

Hands-On Time 🎉

Feedback

THANK YOU!

- **www.data-max.io**
- bujar@data-max.io