ML-Webinar-01

Conventional Machine Learning

Presented By: Vishal Rangras

Moderator: Nehal Soni

Introduction

What is this Webinar about?

- Quick introduction to conventional machine learning
- Dataset exploration, visualization and inference
- Classification and Regression tasks in context
- Path and resources to self-study machine learning

GitHub Repository

All the code and this presentation for this webinar is hosted on the following GitHub repository. Further instructions are provided in the Readme.md of the repository.

URL: ML-Webinars GitHub Repo

Survey

Please fill up the Google Forms survey to provide your valuable feedback about this session

URL: Feedback Form

Libraries and Frameworks

- **Anaconda and Jupyter notebook** - Python, Conda Environment, Notebook Cells.
- **Numpy, Matplotlib, Pandas** - numerical operations, plotting/visualization, columnar data operations
- **Scikit-Learn** - Ready-to-use Machine Learning Algorithms
- **Tensorflow, Keras, PyTorch** - Deep Learning Frameworks
- **Ray, Ray-RLLib** - CPU and GPY Cluster scaling of process, Deep Reinforcement Learning framework
- **OpenCV** - Image Processing and Computer Vision

Resources

YouTube Channels

- **StatQuest** - For Statistics and algorithmic explanation of ML
- **3Blue1Brown** - Essence of linear algebra, Neural Network playlist - Good to visualize math
- **DeepLizard** - For Deep Learning and Deep Reinforcement Learning
- **DeepMind** - Search for David Silver lectures on Reinforcement Learning

Free Courses/Sites

- **Coursera and Udacity** - Lots of free courses from basics to advance level
- **Fast.ai** - Practical Deep Learning
- **MIT OCW** - Open Courseware from MIT for Graduate Level Courses
- **OpenAI and Anyscale Academy** - Lots of resources and tutorials in Deep RL

Machine Learning - 1

Conventional Programming:

- The programmer understands the business logic and code it in programming language.
- It is possible to code business logic / business rules

Machine Learning:

- Useful when it is **hard to code business logic** manually
- Business rules cannot be determined easily
- The idea is to **code learning algorithms** which can **learn** the business logic from **data**
- Useful only with good amount of **useful (correct) data**
- Data and models (algorithms), both are equally important to understand
- **Statistics** is huge part of ML - Khan Academy

Machine Learning - 2

Basic Categorization

- **Supervised Learning** - Ground Truth is known
- **Unsupervised Learning** - No ground truth, but clustering, collaborative filtering, anamoly detection, etc.
- **Reinforcement Learning** - Learning from experience (trial and error)

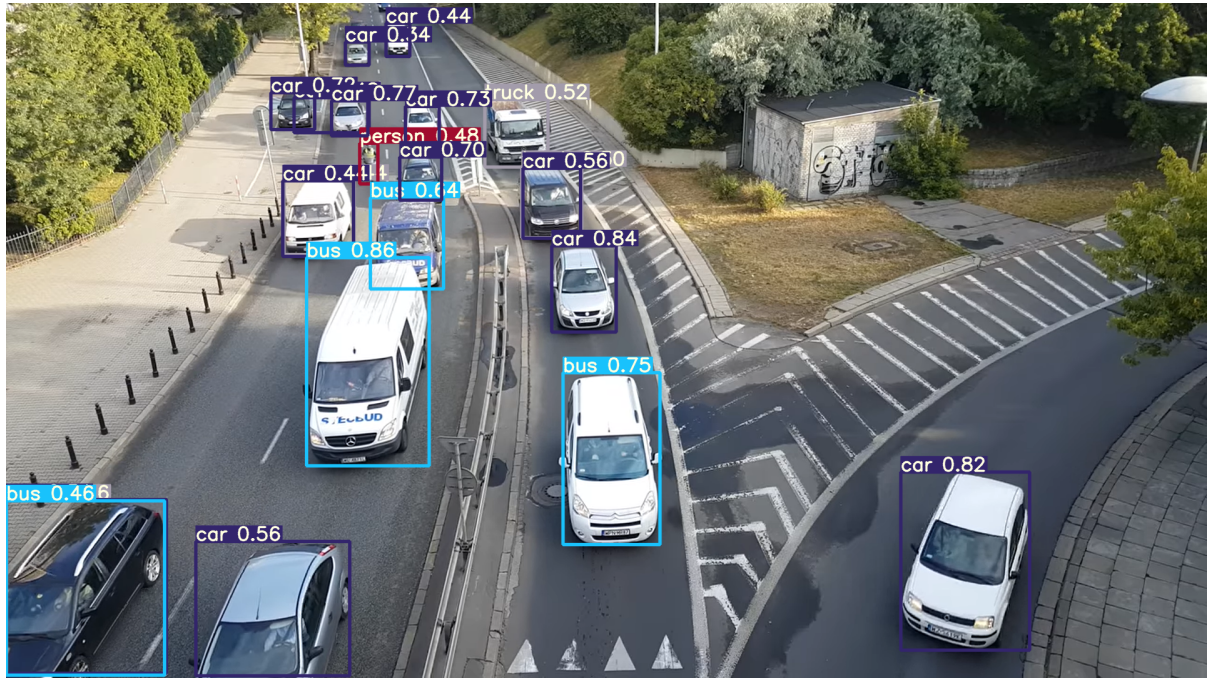
Tasks in Supervised Learning

- **Classification** - Predict discrete value output / Predict from fixed number of categories
- **Regression** - Predict continuous value output

Classification - Example

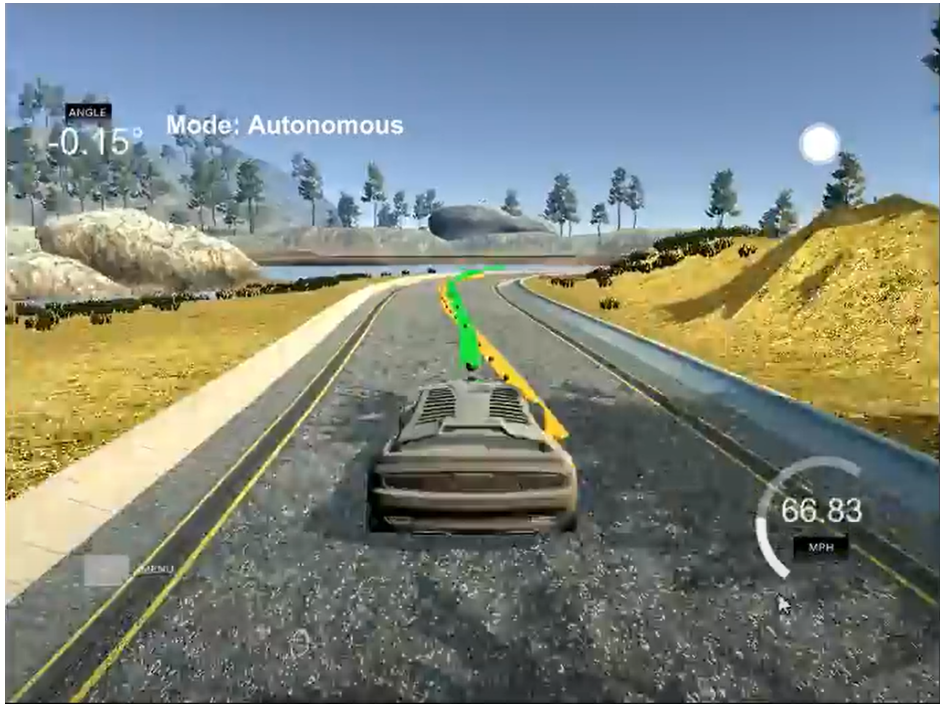
Yolo based Traffic Detection and Tracking

Detections: 1 person, 13 cars, 4 buses, 4 trucks.



Regression - Example

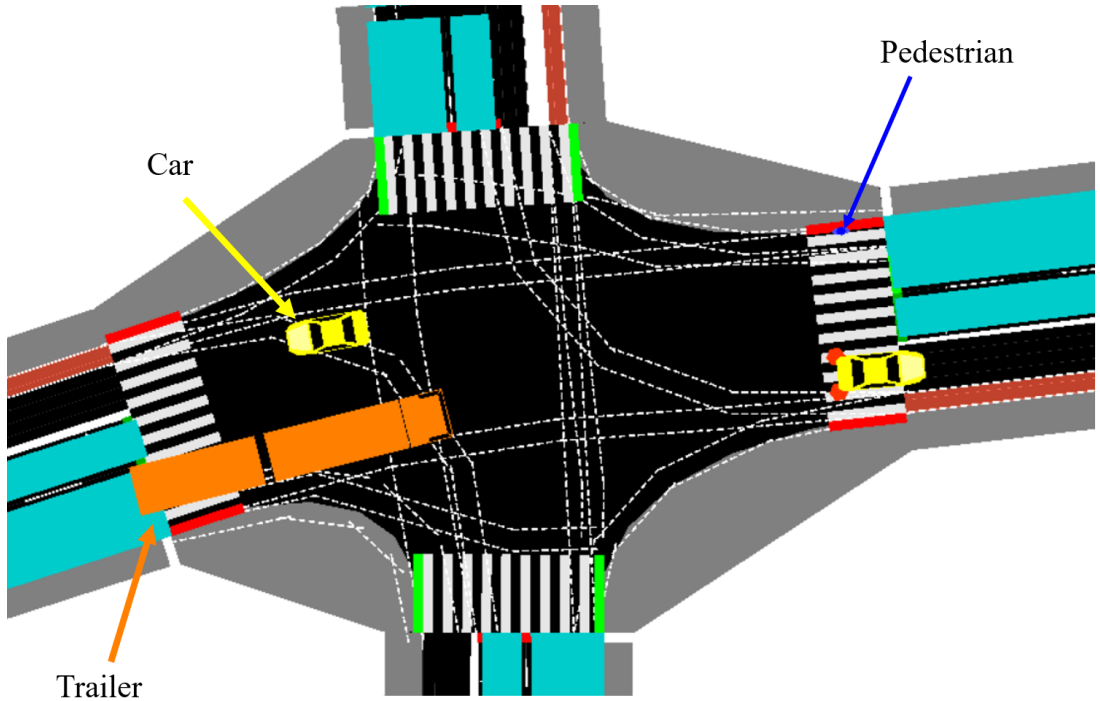
Nvidia's End-to-end Behavioral Cloning



Deep Reinforcement Learning - Example

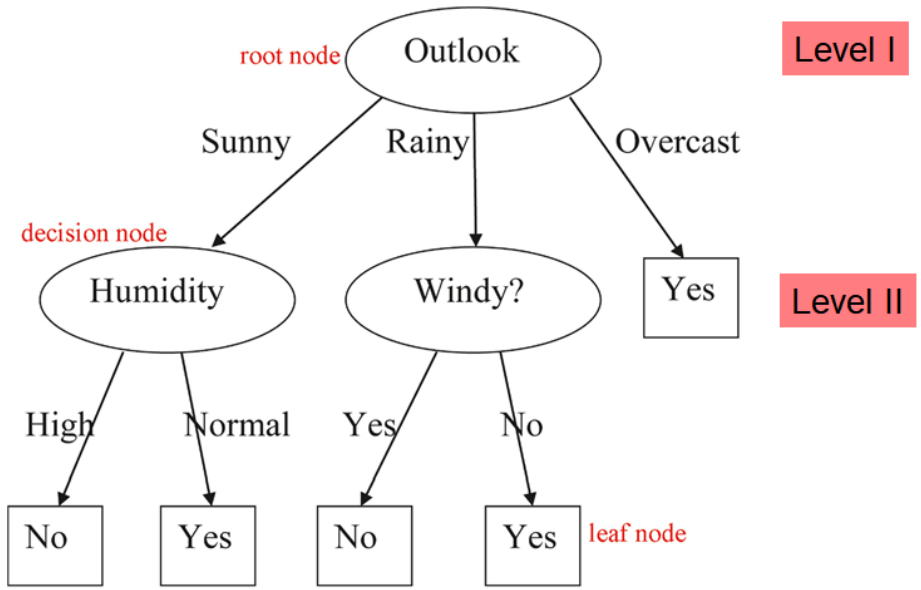
LemgoRL (Courtesy: Fraunhofer IOSB-INA, Lemgo | RL-INA)

GitHub: [LemgoRL Code Repo]



Decision Tree

- Two-level decision tree for determining whether to play tennis
- Non-metric method (non-numerical data)



© book: Dougherty, G. Pattern Recognition and Classification. Springer, 2013

Decision Tree - Code Example 1

Iris Dataset

- **Features / Attributes**: Sepal Length, Sepal Width, Petal Length, Petal Width - (in cm)
- **Labels / Target**: Iris-Setosa, Iris-Versicolour, Iris-Virginica
- **Datapoints**: 150 Samples (50 in each class)
- **Columns** = Features + Labels
- **Rows** = Datapoints

Decision Tree - Code Example 2

Online Shoppers Purchasing Intention Dataset Data Set

- **Data set provided by:** Sakar, C.O., Polat, S.O., Katircioglu, M. et al. Neural Comput & Applic (2018)
- **URL:** Link to UCI Repository

References

1. Yolo v5 for Object Detection
2. Nvidia End-to-end Learning for Self-Driving Cars
3. LemgoRL: An open-source Benchmark Tool to Train Reinforcement Learning Agents for Traffic Signal Control in a real-world simulation scenario
4. Iris Flower Dataset
5. Online Shoppers Purchasing Intent Dataset
6. Python Datascience Handbook