MPSSDEGB
कौशल से रोजगार, समृद्धि अपार

FITT
Foundation For Innovation And Technology Transfer

SANKALP
Ministry of Skill Development
& Entrepreneurship

PRACTICAL AI DEVELOPMENT

DAY 13 HOMEWORK

1. Develop a colab notebook that deals with the some of the functions related to Strings

2. Develop a colab notebook to develop a machine learning model using Numpy, Pandas libraries and present the results with the help of plots using Matplotlib.

OVERVIEW

Setting up an AI development project

Building a small AI model with Tensorflow & Pytorch

Troubleshooting common issues in AI development

Setting up an AI Development Project

To build AI software follow these 7 steps:

1. Identify the problem you want to solve with AI.
2. Gather data.
3. Clean and prepare the data for model training.
4. Choose an AI technology you want to use.
5. Build and train the model.
6. Test the model.
7. Deploy the model.

Setting up an AI development project

1. Identify the problem you want to solve with AI.

- The first thing to do in the AI software development journey is to identify the specific problem you aim to solve.
- This isn't just about having a groundbreaking idea; it's about understanding the nuances of the problem and how AI can solve it.
- This step determines the type of AI technology best suited for your project, whether it's advanced machine learning algorithms, sophisticated natural language processing, or cutting-edge computer vision.

2. Gather data

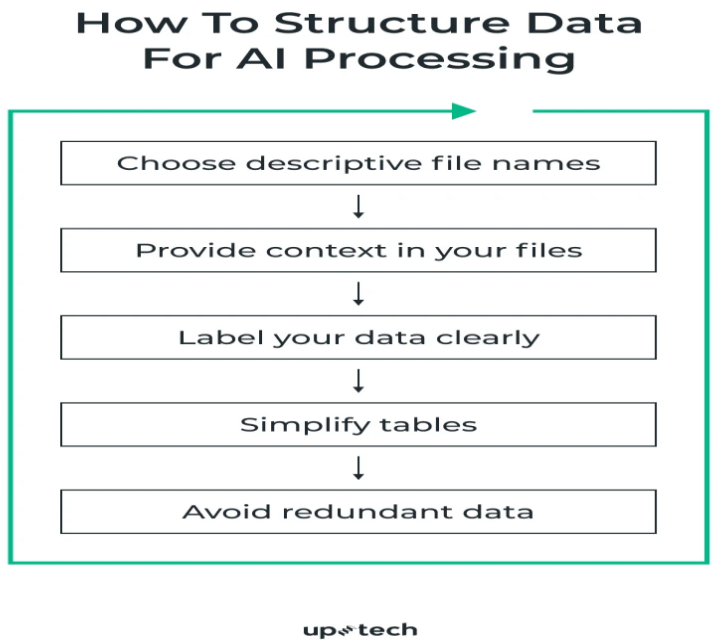
- The next step of making an AI tool should be data gathering.
- It’s data that is relevant to the problem you're addressing, comprehensive enough to cover all possible outcomes, and free from biases.
- Most of the data you'll encounter, especially in AI projects, is unstructured.
- It often requires significant preparation, known as data cleaning.

Structured VS Unstructured Data

Industry	Structured Data	Unstructured Data
E-commerce	• Product IDs • Pricing data	• Customer behaviour and spending patterns • Customer service satisfaction (reviews social media mentions)
Healthcare	• Patient cards • Medical insurance data • Medical billing data	• X-Rays and MRI scans • Treatment recommendations
Banking	• Financial transactions • Customer account data	• Call logs and weblogs • Audio and video communication

3. Clean and prepare the data for model training

- Data cleaning is step 3 in the journey named “How to build an AI tool”.
- This process involves the following parts:
  - ❖ organizing the data;
  - ❖ removing incomplete entries;
  - ❖ classifying it to make it usable for AI training.



4. Choose An AI Technology

Once you have the data, you need to choose the AI technology that best suits your needs.
There are several AI technologies to choose from, such as machine learning, speech recognition, natural language processing, machine learning, augmented reality, and many more.

5. Build And Train The Model

After choosing the AI technology, you need to build and train the model using the gathered data.
This is a complex process that requires expertise in AI and data science.
can create your own AI model without writing a code, and there are a couple of straightforward ways to do this. One of them is using a no-code AI platform. You just provide your data, and the platform handles the training of your AI model.

6. Test The Model

Once the model is built and trained, you need to test it to ensure that it's accurate and reliable.

7. Deploy The Model

Finally, when all the steps are passed you are ready to deploy the model in a production environment where it can be used to solve the users' problem.

PYTORCH:

- PyTorch is a relatively new deep learning framework based on Torch. Developed by Facebook's AI research group and open-sourced on GitHub in 2017, it's used for natural language processing applications.
- PyTorch has a reputation for simplicity, ease of use, flexibility, efficient memory usage, and dynamic computational graphs.
- It also feels native, making coding more manageable and increasing processing speed.

Platforms, Os, Language and other prerequisites					
PyTorch Build	Suitable (1.1)		Preview		
Os	Linux		Mac		Windows
package	Conda	pip	LibTorch	Source	
language	Python 2.7	Python 3.5	Python 3.6	Python 3.7	C++
Cuda	9.0	10.0		None	

In PyTorch, modules are used to represent neural networks.

- Autograd

The autograd module is PyTorch's automatic differentiation engine that helps to compute the gradients in the forward pass in quick time. Autograd generates a directed acyclic graph where the leaves are the input tensors while the roots are the output tensors.

- Optim

The Optim module is a package with pre-written algorithms for optimizers that can be used to build neural networks.

- nn

The nn module includes various classes that help to build neural network models. All modules in PyTorch subclass the nn module.

Pytorch Installation Using pip

Prerequisites
Windows 10 or higher (recommended)
Python 3.8 or higher
CUDA for GPU support (optional)

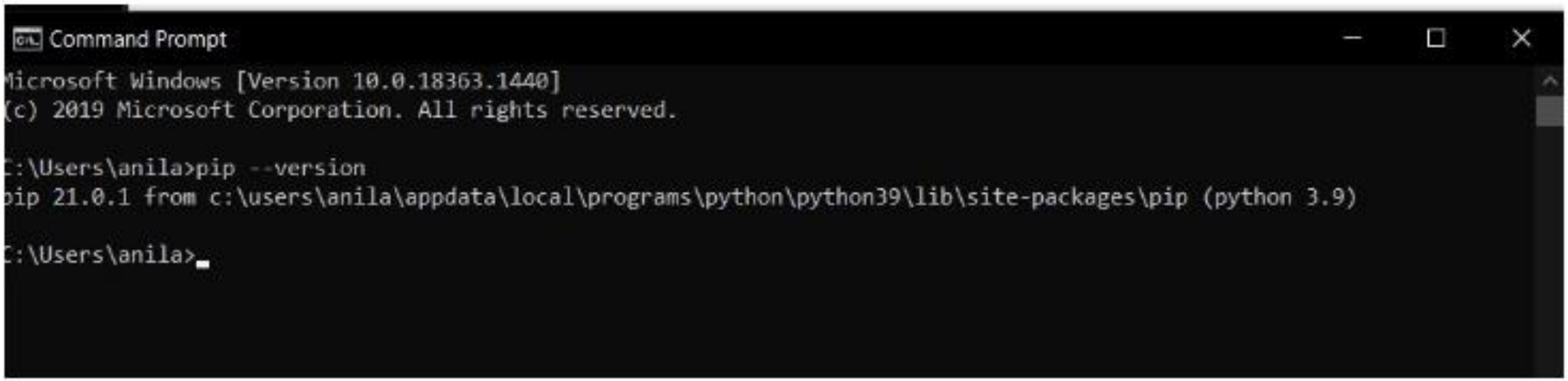
Step 1: Check if python is already installed by entering the following command in the command prompt
.

```
python –version
```

If this command runs successfully, and we are able to get a Python version then we are good to go

Step 2: Check if pip is already installed by entering the following command in the command prompt.

```
pip –version
```



If this command runs successfully, and we are able to get a pip version then we are good to go

Step 3: Enter any one of the following commands *(according to your system specifications)* to install the latest stable release of Pytorch.

1. Compute Platform: CUDA 10.2, Nvidia Driver version should be >= 441.22

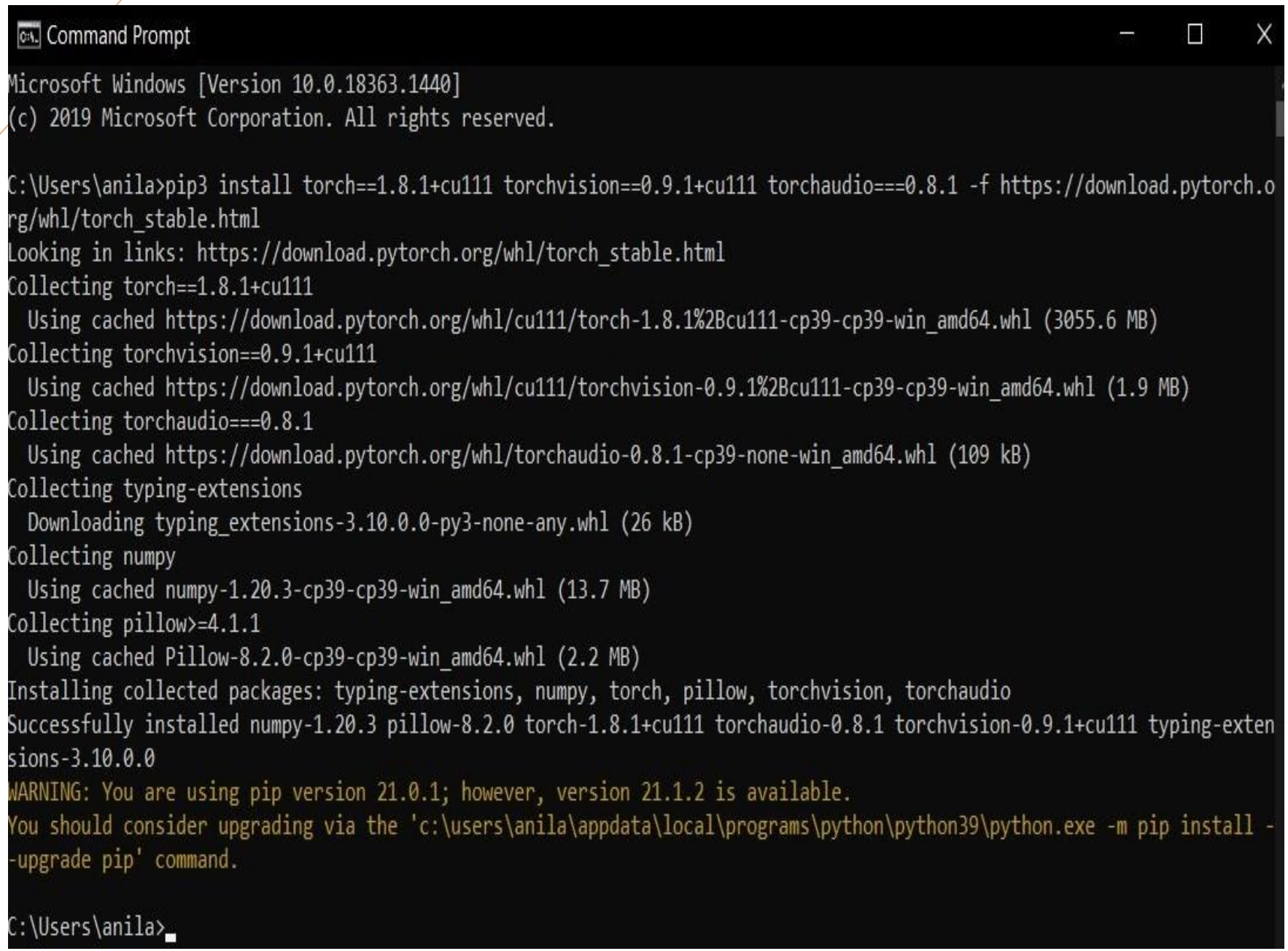
```
pip3 install torch==1.8.1+cu102 torchvision==0.9.1+cu102 torchaudio===0.8.1 -f
https://download.pytorch.org/whl/torch_stable.html
```

2. Compute Platform: CUDA 11.1, Nvidia Driver version should be >= 456.38

```
pip3 install torch==1.8.1+cu111 torchvision==0.9.1+cu111 torchaudio===0.8.1 -f
https://download.pytorch.org/whl/torch_stable.html
```

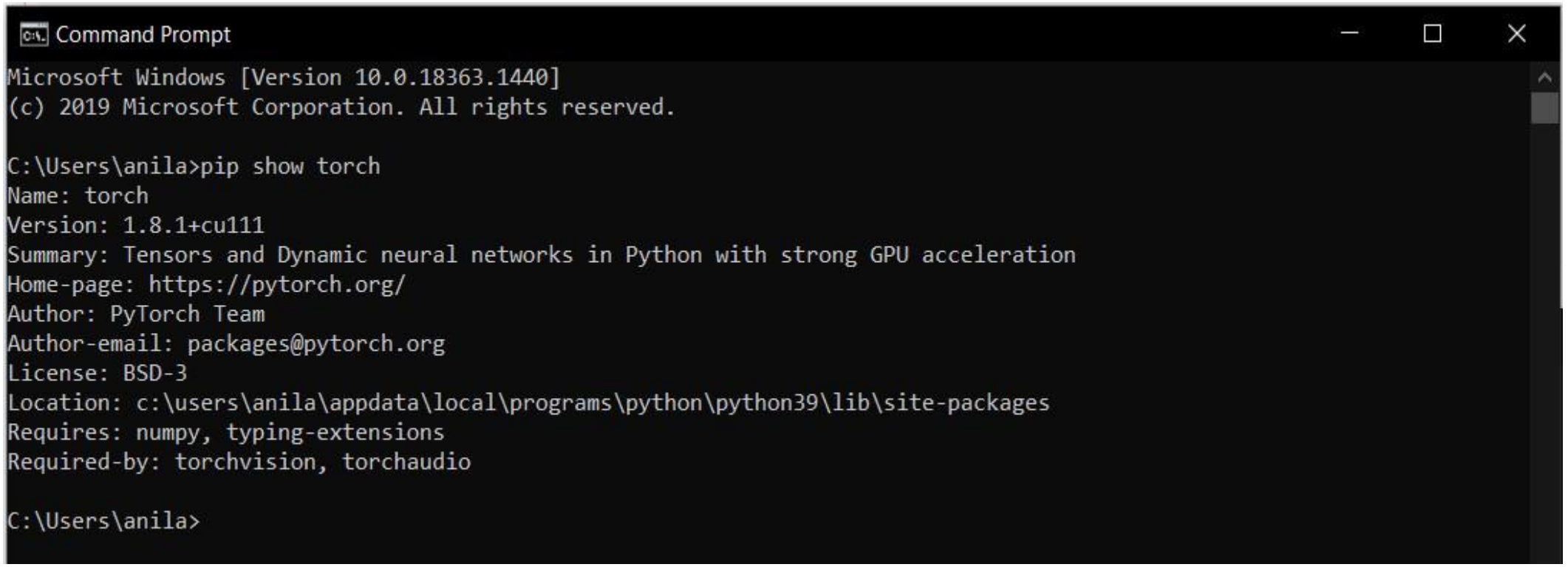
3. Compute Platform: CPU

```
pip3 install torch==1.8.1+cpu torchvision==0.9.1+cpu torchaudio===0.8.1 -f
https://download.pytorch.org/whl/torch_stable.html
```


Command Prompt
Microsoft Windows [Version 10.0.18363.1440]
(c) 2019 Microsoft Corporation. All rights reserved.
C:\Users\anila>pip3 install torch==1.8.1+cu111 torchvision==0.9.1+cu111 torchaudio===0.8.1 -f https://download.pytorch.org/whl/torch_stable.html
Looking in links: https://download.pytorch.org/whl/torch_stable.html
Collecting torch==1.8.1+cu111
Using cached https://download.pytorch.org/whl/cu111/torch-1.8.1%2Bcu111-cp39-cp39-win_amd64.whl (3055.6 MB)
Collecting torchvision==0.9.1+cu111
Using cached https://download.pytorch.org/whl/cu111/torchvision-0.9.1%2Bcu111-cp39-cp39-win_amd64.whl (1.9 MB)
Collecting torchaudio===0.8.1
Using cached https://download.pytorch.org/whl/torchaudio-0.8.1-cp39-none-win_amd64.whl (109 kB)
Collecting typing-extensions
Downloading typing_extensions-3.10.0.0-py3-none-any.whl (26 kB)
Collecting numpy
Using cached numpy-1.20.3-cp39-cp39-win_amd64.whl (13.7 MB)
Collecting pillow>=4.1.1
Using cached Pillow-8.2.0-cp39-cp39-win_amd64.whl (2.2 MB)
Installing collected packages: typing-extensions, numpy, torch, pillow, torchvision, torchaudio
Successfully installed numpy-1.20.3 pillow-8.2.0 torch-1.8.1+cu111 torchaudio-0.8.1 torchvision-0.9.1+cu111 typing-extensions-3.10.0.0
WARNING: You are using pip version 21.0.1; however, version 21.1.2 is available.
You should consider upgrading via the 'c:\users\anila\appdata\local\programs\python\python39\python.exe -m pip install --upgrade pip' command.
C:\Users\anila>

Step 4: Check if Pytorch is successfully installed by entering the following command in the command prompt.

```
pip3 show torch
```



TensorFlow is a popular framework of machine learning and deep learning. It is a free and open-source library which is released on 9 November 2015 and developed by Google Brain Team. It is entirely based on Python programming language and use for numerical computation and data flow, which makes machine learning faster and easier.

TensorFlow can train and run the deep neural networks for image recognition, handwritten digit classification, recurrent neural network, word embedding, natural language processing, video detection, and many more. TensorFlow is run on multiple CPUs or GPUs and also mobile operating systems.

The word TensorFlow is made by two words, i.e., Tensor and Flow

1.Tensor is a multidimensional array

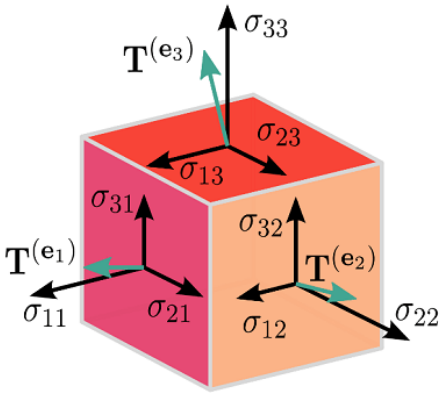
2.Flow is used to define the flow of data in operation.

TensorFlow is used to define the flow of data in operation on a multidimensional array or Tensor.

Components of TensorFlow

Tensor

The name TensorFlow is derived from its core framework, "Tensor." A tensor is a vector or a matrix of n-dimensional that represents all type of data. All values in a tensor hold similar data type with a known shape. The shape of the data is the dimension of the matrix or an array.



In TensorFlow, a tensor is the collection of feature vector (Like, array) of n-dimension. For instance, if we have any 2x3 matrix with values 1 to 6, we write:

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

TensorFlow represents this matrix as:

```
[[1, 3, 5],
[2, 4, 6]]
```

Graphs

TensorFlow makes use of a graph framework. The chart gathers and describes all the computations done during the training.

TensorFlow Installation Using pip

System requirements

- Ubuntu 16.04 or higher (64-bit)
- macOS 10.12.6 (Sierra) or higher (64-bit) *(no GPU support)*
- Windows Native - Windows 7 or higher (64-bit) *(no GPU support after TF 2.10)*
- Windows WSL2 - Windows 10 19044 or higher (64-bit)

Software requirements

- Python 3.9–3.11
- pip version 19.0 or higher for Linux (requires manylinux2014 support) and Windows. pip version 20.3 or higher for macOS.
- Windows Native Requires Microsoft Visual C++ Redistributable for Visual Studio 2015, 2017 and 2019

1. System requirements

- Windows 7 or higher (64-bit)

2. Install Microsoft Visual C++ Redistributable

Go to the Microsoft Visual C++ downloads.
Scroll down the page to the Visual Studio 2015, 2017 and 2019 section.
Download and install the Microsoft Visual C++ Redistributable for Visual Studio 2015, 2017 and 2019 for your platform.

Create a conda environment

Create a new conda environment named tf with the following command.

```
conda create --name tf python=3.9
```

You can deactivate and activate it with the following commands.

```
conda deactivate
conda activate tf
```

Make sure it is activated for the rest of the installation.

Install TensorFlow

TensorFlow requires a recent version of pip, so upgrade your pip installation to be sure you're running the latest version.

```
pip install --upgrade pip
```

If a tensor is returned, you've installed TensorFlow successfully.

PyTorch vs TensorFlow

- Both TensorFlow and PyTorch offer useful abstractions that ease the development of models by reducing boilerplate code.
- They differ because PyTorch has a more "pythonic" approach and is object-oriented, while TensorFlow offers a variety of options.
- PyTorch is used for many deep learning projects today, and its popularity is increasing among AI researchers, although of the three main frameworks, it is the least popular. hen researchers want flexibility, debugging capabilities, and short training duration, they choose PyTorch. It runs on Linux, macOS, and Windows.
- TensorFlow is the favorite tool of many industry professionals and researchers. TensorFlow offers better visualization, which allows developers to debug better and track the training process. PyTorch, however, provides only limited visualization.
- TensorFlow also beats PyTorch in deploying trained models to production, thanks to the TensorFlow Serving framework.
- PyTorch offers no such framework, so developers need to use Django or Flask as a back-end server.

Building a small AI model with Tensorflow & PyTorch

Tensorflow Examples

https://colab.research.google.com/drive/104uYaqWy16nt03_Mg7G09mhw2sjg5QhZ?usp=sharing

https://colab.research.google.com/drive/1D9fnnWLGEZzo-9Xht0k6EeoD7KF_lYHE?usp=sharing

https://colab.research.google.com/drive/1_v50BxPPyL012MPmBj_VuhqU6dgEHQcY?usp=sharing

Pytorch Example

https://colab.research.google.com/drive/1kWNDqVJ7dRrcp9shYf_QZn3cUSmD9LgV?usp=sharing

Troubleshooting Common Issues in AI Development

1. Poor Quality of Data

Data plays a significant role in the machine learning process. One of the significant issues that machine learning professionals face is the absence of good quality data.

Unclean and noisy data can make the whole process extremely exhausting.

We don’t want our algorithm to make inaccurate or faulty predictions. Hence the quality of data is essential to enhance the output.

Therefore, we need to ensure that the process of data preprocessing which includes removing outliers, filtering missing values, and removing unwanted features, is done with the utmost level of perfection.

Troubleshooting common issues in AI development

2. Underfitting of Training Data

This process occurs when data is unable to establish an accurate relationship between input and output variables. It simply means trying to fit in undersized jeans. It signifies the data is too simple to establish a precise relationship.

To overcome this issue:

1. Maximize the training time
2. Enhance the complexity of the model
3. Add more features to the data
4. Reduce regular parameters
5. Increasing the training time of model

Troubleshooting common issues in AI development

3. Overfitting of Training Data

Overfitting refers to a artificial intelligence model trained with a massive amount of data that negatively affect its performance. It is like trying to fit in Oversized jeans.

Unfortunately, this is one of the significant issues faced by many professionals. This means that the algorithm is trained with noisy and biased data, which will affect its overall performance.

Let’s understand this with the help of an example. Let’s consider a model trained to differentiate between a cat, a rabbit, a dog, and a tiger.

The training data contains 1000 cats, 1000 dogs, 1000 tigers, and 4000 Rabbits. Then there is a considerable probability that it will identify the cat as a rabbit. In this example, we had a vast amount of data, but it was biased; hence the prediction was negatively affected.

We can tackle this issue by:

Analyzing the data with the utmost level of perfection
Use data augmentation technique
Remove outliers in the training set
Select a model with lesser features

Troubleshooting common issues in AI development

4. Lack of Training Data

The most important task you need to do in the artificial intelligence process is to train the data to achieve an accurate output.

Less amount training data will produce inaccurate or too biased predictions. Let us understand this with the help of an example.

Consider an algorithm similar to training a child. One day you decided to explain to a child how to distinguish between an apple and a watermelon.

You will take an apple and a watermelon and show him the difference between both based on their color, shape, and taste. In this way, soon, he will attain perfection in differentiating between the two.

But on the other hand, an algorithm needs a lot of data to distinguish. For complex problems, it may even require millions of data to be trained.

Therefore we need to ensure that algorithms are trained with sufficient amounts of data.

Troubleshooting common issues in AI development

5. Slow Implementation

This is one of the common issues faced by many professionals.

The artificial intelligence models are highly efficient in providing accurate results, but it takes a tremendous amount of time.

Slow programs, data overload, and excessive requirements usually take a lot of time to provide accurate results.

Further, it requires constant monitoring and maintenance to deliver the best output.

Troubleshooting common issues in AI development

6. Imperfections in the Algorithm When Data Grows

Finally we have found quality data, trained it amazingly, and the predictions are really concise and accurate.

But the model may become useless in the future as data grows.

The best model of the present may become inaccurate in the coming Future and require further rearrangement.

So we need regular monitoring and maintenance to keep the algorithm working.

This is one of the most exhausting issues faced by many professionals.

Practice material and notebooks

Linear Regression using TensorFlow
https://drive.google.com/file/d/1Z6zPYlS1Ip2fqjVd9VA-o1N19TDYf_jn/view?usp=drive_link

PyTorch CNN model
https://drive.google.com/file/d/1GG12c69gkJwc6Ud2AuexUZwory30wCn6/view?usp=drive_link

TensorFlow tutorial
https://drive.google.com/file/d/1SoftTekn91cjPVaJKjTqPdhDKGt7WkRg/view?usp=drive_link

TensorFlow Tutorial 1
https://drive.google.com/file/d/1SoftTekn91cjPVaJKjTqPdhDKGt7WkRg/view?usp=drive_link

Employee attrition data
https://drive.google.com/file/d/1Pdfjio4c2P0yeLP8YQYD3EpmCwpSqSY2/view?usp=drive_link

HMIS report 2018-19
https://drive.google.com/file/d/16sgUqMTmJX9dli225_6wDScQ9Hl3Bdcj/view?usp=drive_link

HMIS report 2019-20
https://drive.google.com/file/d/1GoEHNbwjTgSXKZwB_OmODHW3QsTFQaVZ/view?usp=drive_link

REFERENCES

https://www.datacamp.com/tutorial/tensorflow-tutorial

https://www.researchgate.net/publication/303657108_TensorFlow_A_system_for_large-scale_machine_learning

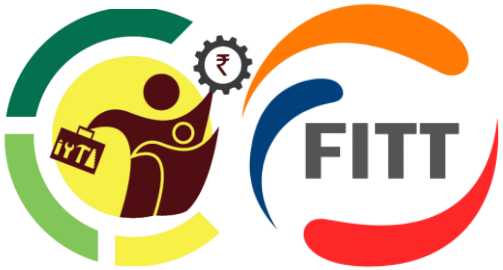
https://www.tensorflow.org/about/case-studies

https://www.simplilearn.com/keras-vs-tensorflow-vs-pytorch-article

HOMEWORK

Develop a basic AI model using Tensorflow

Develop a basic AI model using PyTorch.



THANK YOU