ACES: Accelerating AI/ML Workflows on a Composable Cyberinfrastructure

Zhenhua He Thursday, 27 February 2025 PACES Research Training Workshop





High Performance Research Computing DIVISION OF RESEARCH



AI TechLab

04 Q&A 03 01 (5 mins/lab) 02

Figure 1. Structure of the AI TechLab.

Lab IV. Deep Learning (30 minutes)

We will learn how to use PyTorch to build and train a simple image classification model with deep neural network (DNN).

Lab III Machine Learning (30 minutes)

We will learn to use scikit-learn library for linear regression and classification applications.

2

Lab I. JupyterLab (30 mins)

We will load required modules and activate virtual environment and run JupyterLab on HPRC ACES portal.

Lab II. Data Exploration (30 mins)

We will go through some examples with two popular Python libraries: Pandas and Matplotlib for data exploration.

Ā M

Lab I. JupyterLab



Files	+ 63	± C	E Lorenz.ipynb × E Terminal 1 × Console 1 ×	■ Data.ipynb × ♥ README.md ×							
Ē	♠ > notebooks		B + % □ □ ► ■ C Code ∨	Python 3 C							
	Name 🔺	Last Modified	In this Notebook we explore the Lorenz system	of differential equations:							
2	📃 Data.ipynb	an hour ago		$\dot{x} = \sigma(y - x)$							
	📃 Fasta.ipynb	a day ago		$\dot{y} = \rho x - y - xz$							
_	📃 Julia.ipynb	a day ago		$\dot{z} = -\beta z + xy$							
	 Lorenz.ipynb 	seconds ago									
	R.ipynb	a day ago		. For this set of parameters, we see the trajectories swirling around two points,							
	🖽 iris.csv	a day ago	called attractors.								
	 lightning.json 	9 days ago	In [4]: from lorenz import solve_lorenz								
	nenz.py	3 minutes ago	<pre>t, x_t = solve_lorenz(N=10)</pre>								
			Output View × Internet Intern	x							
				<pre>lve_lorenz(N=10, max_time=4.0, sigma=10.0, beta=8./3, rho=28.0):</pre>							
				<pre>'Plot a solution to the Lorenz differential equations.""" g = plt.figure()</pre>							
				= fig.add_axes([0, 0, 1, 1], projection='3d')							
			rho 28.00 13 ax	axis('off')							
				prepare the axes limits							
				<pre>set_xlim((-25, 25))</pre>							
				.set_ylim((-35, 35)) .set_zlim((5, 55))							
			16 dx.	Set_2thm((5, 55))							
				f lorenz_deriv(x_y_z, t0, sigma=sigma, beta=beta, rho=rho):							
				<pre>"""Compute the time-derivative of a Lorenz system.""" x, y, z = x_y_z</pre>							
			23	return [sigma * (y - x), x * (rho - z) - y, x * y - beta * z]							
			24	Never made an entry and a straight and from 15 to 15							
			25 # 1	Choose random starting points, uniformly distributed from -15 to 15 .random.seed(1)							

NSF ACES

Accelerating Computing for Emerging Sciences

Our Mission:

- NSF ACSS CI testbed
- Offer an accelerator testbed for numerical simulations and AI/ML workloads
- Provide consulting, technical guidance, and training to researchers
- Collaborate on computational and data-enabled research.



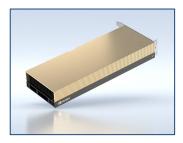
ACES Accelerators/Components

Component	Quantity	Description
Graphcore IPU	32	16 Colossus GC200 IPUs, 16 Bow IPUs. Each IPU group hosted with a CPU server as a POD16 on a 100 GbE RoCE fabric
FPGAs:		
Intel PAC D5005	2	Accelerator with Intel Stratix 10 GX FPGA and 32 GB DDR4
BittWare IA-840F	3	Accelerator with Agilex AGF027 FPGA and 64 GB of DDR4
NextSilicon Coprocessor	2	Reconfigurable accelerator with an optimizer continuously evaluating application behavior.
NEC Vector Engine	8	Vector computing card (8 cores and HBM2 memory)
Intel Optane SSD	48	18 TB of SSDs addressable as memory w/ MemVerge Memory Machine.
NVIDIA GPUs:		
H100	30	For HPC, DL Training, Al Inference
A30	4	For AI Inference and Mainstream Compute
Intel PVC GPUs	120	Intel GPUs for HPC, DL Training, Al Inference

Refer to our **Knowledge Base** for more: <u>https://hprc.tamu.edu/kb/User-Guides/ACES/Hardware/</u>

Ā M

NVIDIA GPUS on ACES



A30s support less intense workloads relying on numerical simulations and AI/ML methods.



H100s Support computationally intensive workloads employing numerical simulations and AI/ML methods.

Specify in Slurm file:
 #SBATCH --partition=gpu
 #SBATCH --gres=gpu:<gpu type>:<number>

Intel Data Center Max GPU 1100 GPUs (PVC GPUs)



PVC training event (<u>Link</u>) planned for March 4th!

See also our Youtube channel for previous PVC training sessions!

Intel GPUs for HPC, DL Training, AI Inference

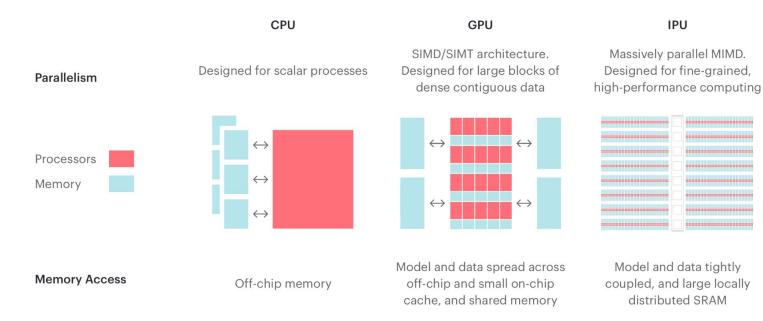
Specify in Slurm file:

#SBATCH --partition=pvc
#SBATCH --gres=gpu:pvc:<number>

See our command-line tool to check configuration:

\$ show_	pvc_features		
HOSTNAME	AVAIL_FEATURES	GRES	STATE
ac010	gen4_fabric	gpu:pvc:4	mixed
ac024	gen4_fabric	gpu:pvc:8	idle
ac025	gen4_fabric	gpu:pvc:4	mixed
ac026	gen5_fabric	gpu:pvc:6	reserved
ac030	gen5_fabric	gpu:pvc:8	reserved
ac081	gen5_fabric,xelink4	gpu:pvc:4	reserved
ac082	gen5_fabric,xelink2	gpu:pvc:2	reserved

Graphcore IPUs



www.graphcore.ai/bow-processors

Ă_M ⊦

High Performance Research Computing | hprc.tamu.edu | NSF Award #2112356

8

Graphcore IPUs

GRAPHCORE

Porting TensorFlow 2 Models Quick Start Version: Latest



1. Import the TensorFlow IPU module

2. IPU Config

3. Model

4. Training process

5. Optimization

6. Trademarks & copyright



🕒 Download PDF

I. IMPORT THE TENSORFLOW IPU MODULE

First, we import the TensorFlow IPU module.

Add the import statement in Listing 1.1 to the beginning of your script.

Listing 1.1 Importing ipu Python module

from tensorflow.python import ipu

For the *ipu* module to function properly, we must import it directly rather than accessing it through the top-level TensorFlow module.

2. IPU CONFIG

To use the IPU, you must create an IPU session configuration in the main process. A minimum configuration is in Listing 2.1.

Listing 2.1 Example of a minimum configuration

ipu_config = ipu.config.IPUConfig()
ipu_config.auto_select_ipus = 1 # Select 1 IPU for the model
ipu_config.configure_ipu_system()

This is all we need to get a small model up and running. A full list of available in the Python API documentation.

docs.graphcore.ai/en/latest

Accessing Graphcore IPUs

- SSH into poplar1 or poplar2 from ACES
 - o [username@login ~]\$ ssh poplar1 ---
- Enable the SDK environment. See our KB for details:
 - <u>hprc.tamu.edu/kb/User-Guides/ACES/Graphcore_Colossus_IPUs/</u>
 - <u>hprc.tamu.edu/kb/User-Guides/ACES/Graphcore_Bow_IPUs/</u>
- Type gc-monitor to view the status of the IPUs:

gc-monitor	Partition: p16 [active] has 16 reconfigurable IPUs										
IPU-M	Serial	IPU-M SW	Server version	ICU FW	Type	ID	IPU#	Routing			
10.1.5.1	0010.0002.8213921	1	1.9.0	2.4.4	M2000	0	3	DNC			
10.1.5.1	0010.0002.8213921	i i	1.9.0	2.4.4	M2000	11	2	DNC			
10.1.5.1	0010.0001.8213921	i i	1.9.0	2.4.4	M2000	2	1	DNC			
10.1.5.1	0010.0001.8213921	į į	1.9.0	2.4.4	M2000	3	0	DNC			
10.1.5.2	0030.0002.8213921	1	1.9.0	2.4.4	M2000	4	3	DNC			
10.1.5.2	0030.0002.8213921	i i	1.9.0	2.4.4	M2000	5	2	DNC			
10.1.5.2	0030.0001.8213921	i i	1.9.0 1	2.4.4	1 M2000	6	i 1	I DNC			

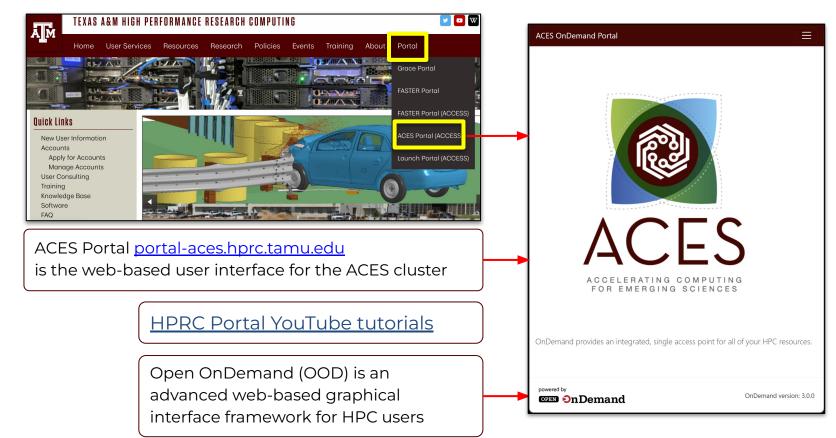
IPU training event (<u>Link</u>) planned for March 18th!

Contact us first to be given access to poplar

help@hprc.tamu.edu

See also our Youtube channel for previous IPU training sessions!

ACES Portal



High Performance Research Computing | hprc.tamu.edu | NSF Award #2112356

11

Accessing ACES via the Portal (ACCESS)

LOCATIONS SUPPORT OPERATIONS METRICS A C E Login	
ACCESS	If you had an XSEDE account, please enter your XSEDI username and password for ACCESS login. ACCESS ID
consent to Attribute Release	
TAMU ACES ACCESS OIDC requests access to the following information. If you do not approve this request, do not proceed. • Your ClLogon user identifier • Your name • Your email address	ACCESS Password
Your username and affiliation from your identity provider	LOGIN
ct an Identity Provider	
Remember this selection 0	Register for an ACCESS ID Forgot your password?
	Need Help?
Select the Identity Provider appropriate for your account.	Log in using your ACCES

ĂМ

institutional credentials.

Get a Shell on ACES

Click on "Clusters" menu → _aces Shell Access



Success!

Welcome to the ACES login node.

Check which login node you are on.

АМ

Host: login.aces				Themes:	Default				
Consulting: ACES Documentation: FASTER Documentation: Grace Documentation: Terra Documentation: YouTube Channel:	help@hprc.tamu.edu https://hprc.tamu.e https://hprc.tamu.e https://hprc.tamu.e https://hprc.tamu.e https://www.youtube	du/kb/Use du/kb/Use du/kb/Use du/kb/Use .com/texa	r-Guides/ACES r-Guides/FAS r-Guides/Grad r-Guides/Tern samhprc	S [TER] ce]					
****	*****	******	*****	*******					
* === I	MPORTANT POLICY INFO	RMATION =		>	k				
* - Unauthorized use of I		ohibited	and subject †	to >	k				
* criminal prosecution					k				
* - Use of HPRC resource					k 				
 * laws and regulations * US citizens and legal 			Starr members		k k				
* - Sharing HPRC account		nation is	in violation		k				
* Texas State Law. Any					k				
* - Authorized users must				>	k				
	tps://hprc.tamu.edu/				k				
*******	*****	*******	******	******					
!! WARNING: THERE ARE	ONLY NIGHTLY BACKUP	S OF USER	HOME DIRECTO	ORIES. !!					
Please restrict usage to <u>8 CORES</u> across ALL login nodes. Users found in violation of this policy will be <u>SUSPENDED</u> .									
To see these	messages again, run	the motd	command.						
our current disk quotas a									
isk	Disk Usage	Limit	File Usage	Limit					
home/u.zh108696	4.0G	10.0G	2361	10000					
scratch/user/u.zh108696 ype 'showquota' to view t u.zh108696 <mark>@aces-login1</mark> ~]		1.0T	352057	1000000					

Commands to copy the materials

• Navigate to your personal scratch directory

\$ cd \$SCRATCH

• Files for this course are located at

/scratch/training/paces_Feb2025

Make a copy in your personal scratch directory

Please skip the first two steps if you have done them.

\$cp -r /scratch/training/paces_Feb2025 \$SCRATCH

• Enter this directory (your local copy)

\$cd paces Feb2025/ai tech labs

Go to JupyterLab Page

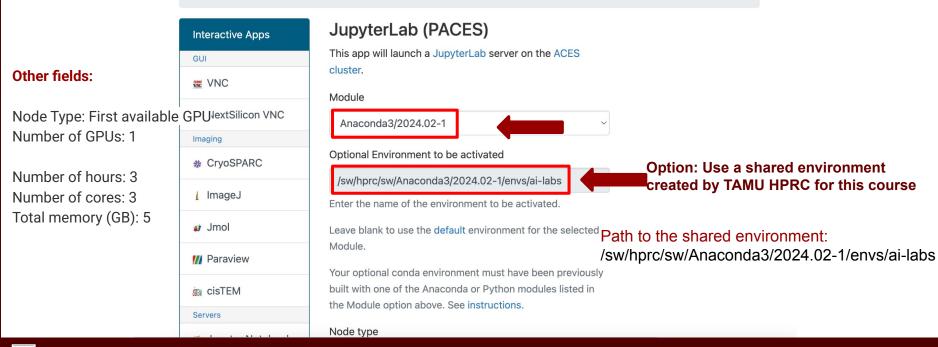
ACES OnDemand Portal Files - Jobs - Clusters -	Interactive Apps 🝷 Affinit	y Groups 👻 Chatbot 👻 Dashboard 👻
	GUI VNC NextSilicon VNC	
OnDemand provides an integrated, single acces	Imaging & CryoSPARC ImageJ Jmol Paraview a cisTEM	resources.
Message of the Day IMPORTANT POLICY INFORMATION • Unauthorized use of HPRC resources is prohibite • Use of HPRC resources in violation of United Star • Sharing HPRC account and password information DISABLED.	 Servers ⇒ Jupyter Notebook ⇒ JupyterLab ⇒ JupyterLab (PACES) ● RStudio 	osecution. egulations is prohibited. All control accounts will be
Authorized users must also adhere to ALL policie	✤ TensorBoard	policies

!! WARNING: THERE ARE ONLY NIGHTLY BACKUPS OF L

JupyterLab Page

ACES OnDemand Portal Files
Jobs Clusters Interactive Apps Affinity Groups Chatbot Dashboard

Home / My Interactive Sessions / JupyterLab (PACES)



High Performance Research Computing | hprc.tamu.edu | NSF Award #2112356

17

A M

Connect to JupyterLab

Ā M

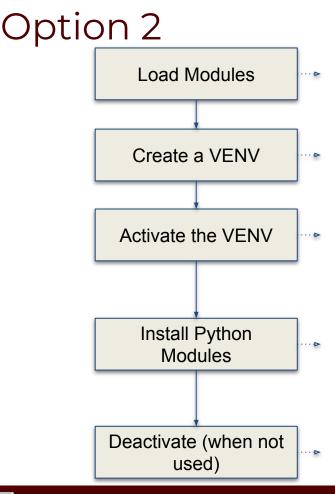
ACES OnDemand Portal Files 🔻 Interactive Apps -Affinity Groups 🔻 Chatbot -Dashboard -Utilities • Jobs 🔻 Clusters -Session was successfully created. X Home / My Interactive Sessions Interactive Apps JupyterLab (PACES) (793933) 1 node | 3 cores | Running GUI Host: >_ ac035 8 Delete VNC VNC Created at: 2025-02-26 10:23:31 CST **WEXTSILICON VNC** Time Remaining: 59 minutes Imaging Session ID: 309b30c3-f959-40a6-b860-47b925b59388 CryoSPARC Connect to JupyterLab ImageJ

Review and Exercise

- Log into ACES through ACES Portal (ACCESS)
- Copy the training materials to your \$SCRATCH directory
- Launch JupyterLab app
- In the notebook named *01_Jupyterlab.ipynb*, follow the instructions to import the

required modules to make sure they have been loaded properly.

⊠L	.auno	cher				×		01_J	upyterlab.ipynb	×	+		
8	+	Ж	ē	Ċ	•		C	**	Markdown 🗸			Python 3 (ipykernel)	0
	[1:							and click [Shif _ TechLab!")	t+Ente	er]	to execute	
	[1:	# to # w	est r rite			de b	elow	/				
			► C	lick ł	nere	to s	ee s	olut	tion				
	[1:	# to # w	est µ rite			de b	elow	V				
			•										. 1
			С	lick ł	nere	to s	ee s	olut	tion				



A M

clean up and load Anaconda cd \$SCRATCH module purge module load Anaconda3/2024.02-1

create a Python virtual environment
conda create -n ai-labs

activate the virtual environment
source activate ai-labs

install required package to be used in the portal conda install -c anaconda jupyter conda install -c anaconda pandas conda install -c conda-forge matplotlib conda install -c anaconda scikit-learn conda install -c conda-forge transformers conda install pytorch torchvision torchaudio pytorch-cuda=11.8 -c pytorch -c nvidia

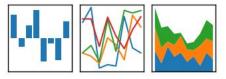
deactivate the virtual environment
source deactivate

Lab II. Data Exploration

matpletlib

Ā M





Data Structures

Pandas has two data structures that are descriptive and

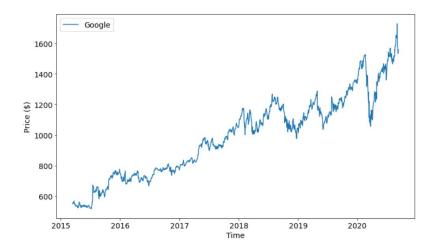
optimized for data with different dimensions.

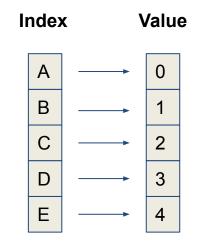
- Series: 1D labeled array
- DataFrame: General 2D labeled, size-mutable tabular

structure with potentially heterogeneously-typed columns

Series in pandas

- One-dimensional labeled array
- Capable of holding any data type (integers, strings, floating point numbers, etc.)
- Example: time-series stock price data



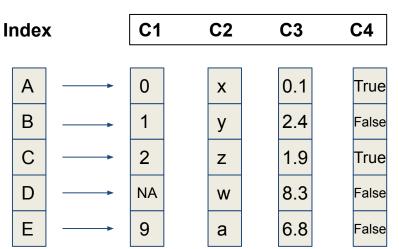


DataFrame in pandas

- Primary Pandas data structure
- A dict-like container for Series objects
- Two-dimensional size-mutable
- Heterogeneous tabular data structure

Α	В	С	D	E	F	G	н
id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors
7129300520	20141013T0	221900	3	1	1180	5650	1
6414100192	20141209T0	538000	3	2.25	2570	7242	2
5631500400	20150225T0	180000	2	1	770	10000	1
2487200875	20141209T0	604000	4	3	1960	5000	1
1954400510	20150218T0	510000	3	2	1680	8080	1
7237550310	20140512T0	1.23E+06	4	4.5	5420	101930	1
1321400060	20140627T0	257500	3	2.25	1715	6819	2
2008000270	20150115T0	291850	3	1.5	1060	9711	1
2414600126	20150415T0	229500	3	1	1780	7470	1

Ā M



Columns

Pandas Learning Objectives

After this lesson, you will know how to:

- Create a DataFrame
- Retrieve a Row or Column
- Drop Entries
- Index, Select, and Filter data
- Sort data

ĂМ

• Input and Output



Key Plotting Concepts in Matplotlib

Matplotlib: Figure

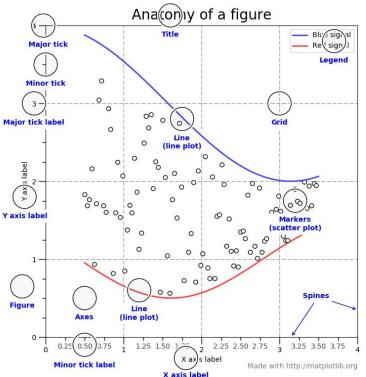
Figure is the object that keeps the whole image output. Adjustable parameters include:

- 1. Image size (set_size_inches())
- 2. Whether to use tight_layout (set_tight_layout())

• Matplotlib: Axes

Axes object represents the pair of axis that contain a single plot (x-axis and y-axis). The Axes object also has more adjustable parameters:

- The plot frame (set_frame_on() or set_frame_off())
- X-axis and Y-axis limits (set_xlim() and set_ylim())
- X-axis and Y-axis Labels (set_xlabel() and set_ylabel())
- 4. The plot title (set_title())



(Credit: matplotlib.org)

Matplotlib Learning Objectives

After this lesson, you will know how to create:

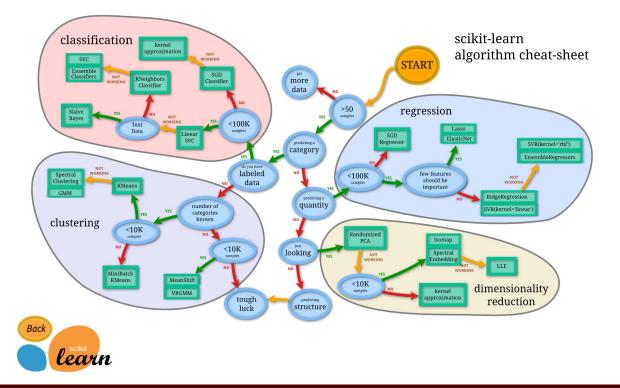
- Scatter plot and Line plot
- Subplots
- Color map
- Contour figures
- 3D figures

ĂМ

- Surface plots
- Wire-frame plot
- Contour plots with projections

JupyterLab Exercises

Lab III. Machine Learning



Main Features of scikit-learn



Classification	Regression	Clustering	Dimension Reduction	Model Selection	Preprocessing
Identifying category of an object	Predicting a attribute for an object	Grouping similar objects into sets	Reducing the number of dimensions	Selecting models with parameter search	Preprocessing data to prepare for modeling
Applications: Spam detection, image recognition. Algorithms: SVM, nearest neighbors, random forest, and more	Applications: Drug response, Stock prices. Algorithms: SVR, nearest neighbors, random forest, and more	Applications: Customer segmentation, Grouping experiment outcomes Algorithms: k-Means, spectral clustering, mean-shift, and more	Applications: Visualization, Increased efficiency Algorithms: k-Means, feature selection, non-negative matrix factorization, and more	Applications: Improved accuracy via parameter tuning Algorithms: grid search, cross validation, metrics, and more	Applications: Transforming input data such as text for use with machine learning algorithms. Algorithms: preprocessing, feature extraction, and more
$\overline{}$	E %	°		$\hat{\mathbf{r}}$	᠕᠕

6

JupyterLab Exercises

0

0 \ 0

Lab IV. Deep Learning

Deep Learning

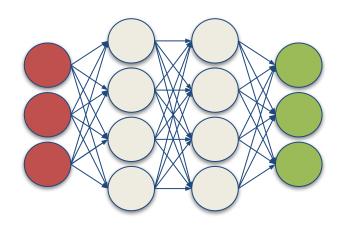
by Ian Goodfellow, Yoshua Bengio, and Aaron Courville http://www.deeplearningbook.org/

Animation of Neutron Networks

by Grant Sanderson <u>https://www.3blue1brown.com/</u>

Visualization of CNN

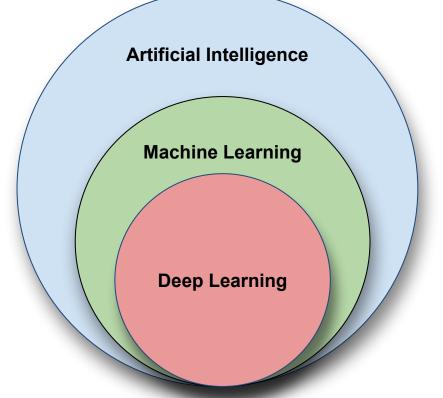
by Adam Harley <u>https://adamharley.com/nn_vis/cnn/3d.html</u>





Relationship of AI, ML, and DL

- Artificial Intelligence (AI) is anything about man-made intelligence exhibited by machines.
- Machine Learning (ML) is an approach to achieve AI.
- **Deep Learning (DL)** is one technique to implement **ML**.



Types of ML Algorithms

• Supervised Learning

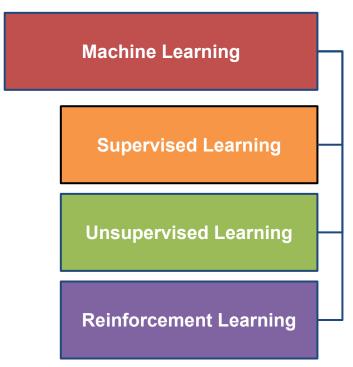
 trained with labeled data; including regression and classification problems

• Unsupervised Learning

 trained with unlabeled data; clustering and association rule learning problems.

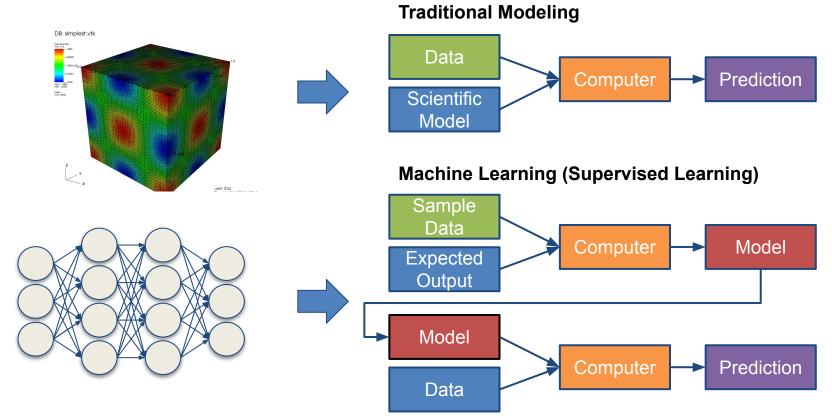
Reinforcement Learning

 no training data; stochastic Markov decision process; robotics and business strategy planning.

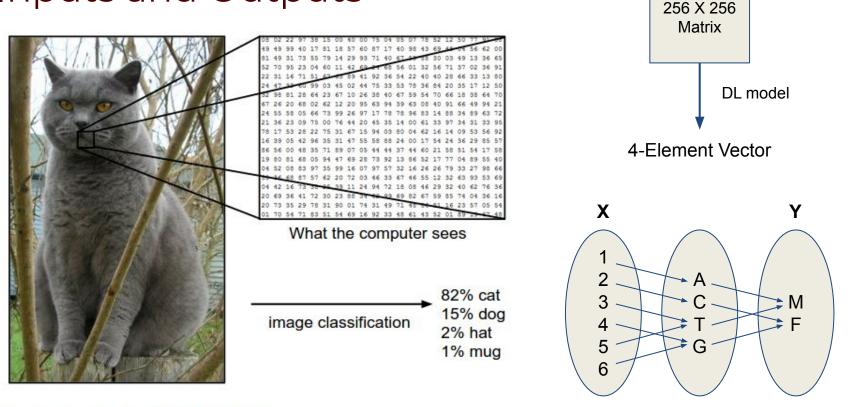


Machine Learning

ĀМ



Inputs and Outputs

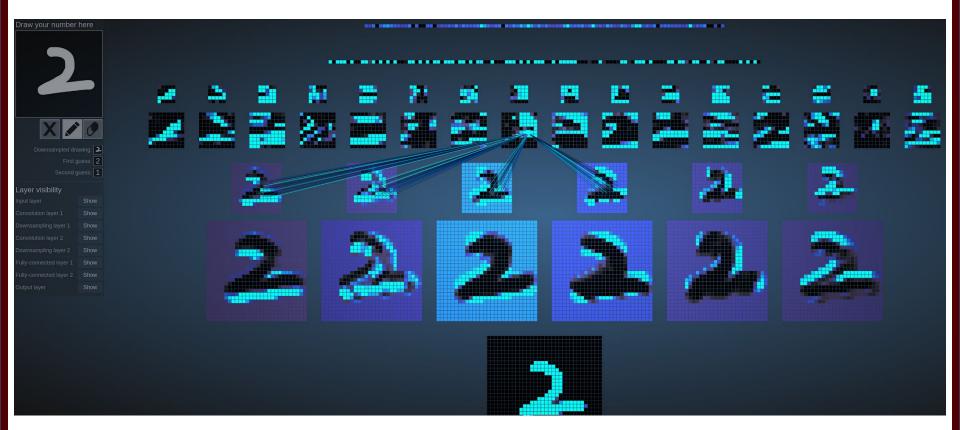


With deep learning, we are searching for a **surjective** (or **onto**) function **f** from a set **X** to a set **Y**.

Image from the Stanford CS231 Course

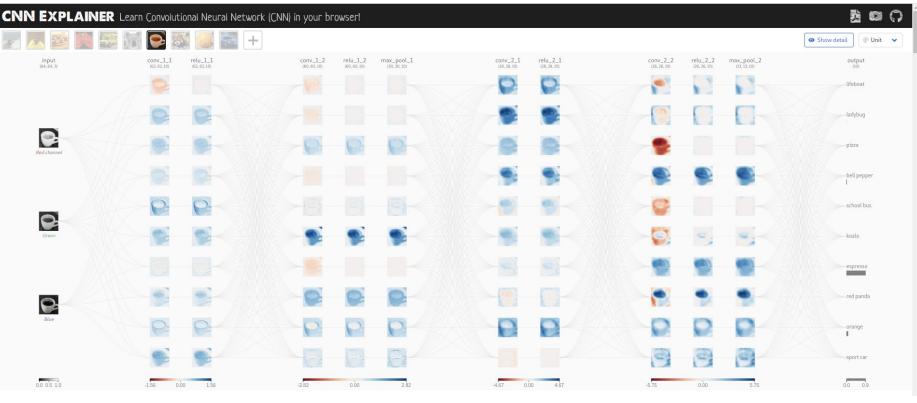
M

MNIST - CNN Visualization



(Image Credit: https://adamharley.com/nn_vis/cnn/3d.html)

CNN Explainer



(Image Credit: https://poloclub.github.io/cnn-explainer/)

ĀМ

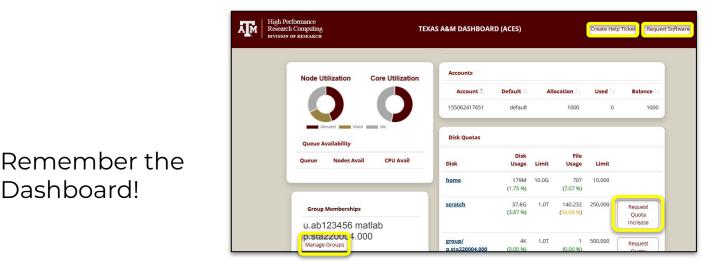
JupyterLab Exercises

Need Help?

First check the FAQ: https://hprc.tamu.edu/kb/FAQ/Accounts

- ACES user Guide: <u>https://hprc.tamu.edu/kb/User-Guides/ACES</u>
- Email your questions to help@hprc.tamu.edu

Dashboard!



Need Help?

Help us help you -- tell us:

- Which cluster
- Username
- Job id(s) if any
- Location of your jobfile, input/output files
- Application used if any
- Module(s) loaded if any
- Error messages
- Steps you have taken, so we can reproduce the problem