IOWA STATE UNIVERSITY

Translational AI Center

Intro to ML/AI with HPC



"The training of programs developed by allowing a computer to learn from its experience, rather than through manually coding the individual steps." fast.ai book

Artificial Intelligence

Large Language Models

Description: Advanced Al systems designed to understand and generate human language

Examples: Transformer-based models such as GPT and LLAMA

Data Type & Size: Large datasets; process and generate text by learning from vast amounts

Complexity: High - These models can understand context, generating coherent text, and performing various language-related tasks

Examples of Usage in Agriculture: Extension services and Crop Management Advice

Generative Al

Description: A class of artificial intelligence techniques that create new content by learning from existing data

Examples: Generative Adversarial Networks and Variational Autoencoders

Data Type & Size: Large datasets designed to produce new, original outputs such as images, text, or sound

Complexity: High - These models learn the underlying patterns and structures in the training data and generate novel data that mimics the characteristics of the original data

Examples of Usage in Agriculture: Synthetic Data Generation for phenotyping tasks

Deep Learning

Description: it is a subset of machine learning and artificial intelligence that focuses on algorithms inspired by the structure and function of the human brain

Examples: Autoencoders, Recurrent Neural Networks, Convolutional Neural Networks, Generative Adversarial Networks

Data Type & Size: Large datasets; images, text, and audio

Complexity: High -Automatically learn and represent data with multiple levels of abstraction, less explainable in comparison with Machine learning methods

Examples of Usage in Agriculture: Crop yield prediction, disease detection

Description: A subset of Al focused on developing

Machine Learning

Al focused on developing algorithms that enable machines to learn patterns from data and make predictions or decisions.

Examples: Decision Trees, Naïve Bayes, K-Nearest Neighbors, and Support Vector Machines (SVMs)

Data Type & Size: Often small to medium-sized datasets; tabular, time series, or image data.

Complexity: Moderate -Generally more explainable than deep learning models.

Examples of Usage in Agriculture: Characterization of soybean nodulation, disease detection Description: Al encompasses various techniques and models that enable machines to perform tasks that typically

Examples: Includes both rule-based systems and learning-based models

require human intelligence

Data Type & Size: Can handle any data type; size varies from small datasets to large-scale data

Complexity: Varies from simple rules to complex deep learning models

Examples of Usage in Agriculture: All examples across all subsets



Machine Learning Workflow

Training Model with random parameters Model Training with trained parameters Data Inference Model Predictions Inputs with trained parameters



What Machine Learning can do?

What can we do with machine learning?

Play chess, go, ..

Classify data: 😭 vs. 😺 , galaxies, 🖺 species from their calls, ...

Recommender systems: 🆀 / 🖳 / 📜 suggestions, ...

Solve $\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = v \frac{\partial^2 u}{\partial x^2}$

Cluster data into different groups

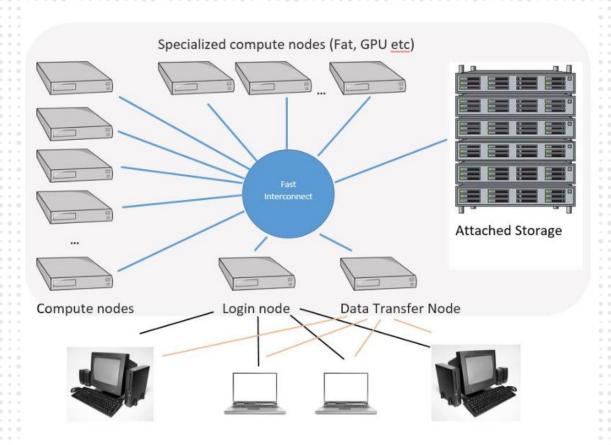
Write poetry, create art



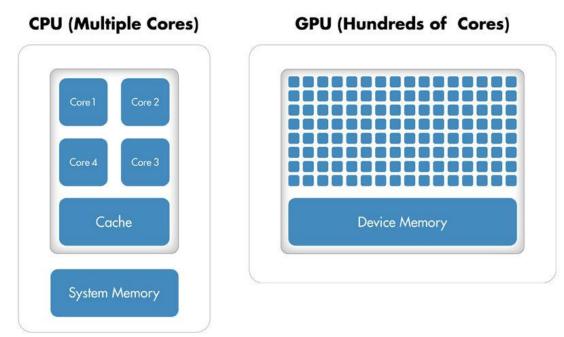
Things to keep in mind

- Computers aren't smart but can be (over)confident
- While we do not manually code the individual steps in the program, we still do a lot
 - We decide which model, data, training hyperparameters, ...
- Data is central to machine learning
 - Untrained models are generally no better than random chance
 - Trained models often learn biases in data

HPC



https://www.hpc.iastate.edu/



https://la.mathworks.com/

AI and HPC: Benefits

Faster Training Times

Larger Model Scalability

Cost-Effective

Improved Accuracy

Larger models can't be trained without multiple GPUs

modifier_ob. mirror object to mirror mirror_mod.mirror_object peration == "MIRROR_X": _____mod.use_x = True urror_mod.use_y = False irror_mod.use_z = False _operation == "MIRROR_Y" irror_mod.use_x = False lrror_mod.use_y = True lrror_mod.use_z = False _operation == "MIRROR_Z": rror_mod.use_x = False rror_mod.use_y = False rror_mod.use_z = True melection at the end -add ob.select= 1 er ob.select=1 ntext.scene.objects.action "Selected" + str(modified rror ob.select = 0 bpy.context.selected_obj ata.objects[one.name].sel int("please select exactle --- OPERATOR CLASSES ---pes.Operator): mirror to the selected ject.mirror_mirror_x" ext.active_object is not

Best Practices for Using HPC for Machine Learning

Choose: The Right HPC System

Optimize: Code for HPC

Use: Parallel Processing

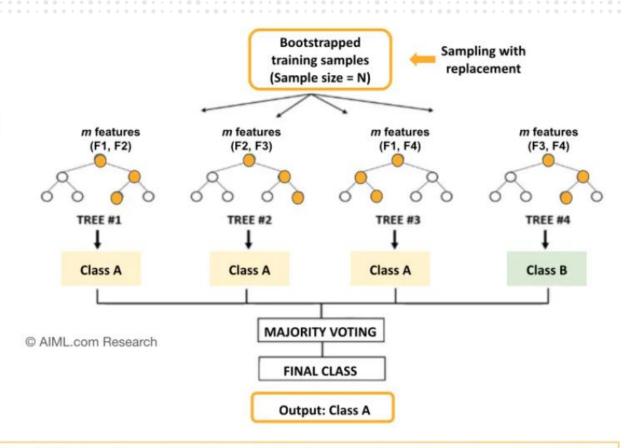
Monitor: Performance

Use: Machine Learning Optimized

Libraries

Random Forest for Classification

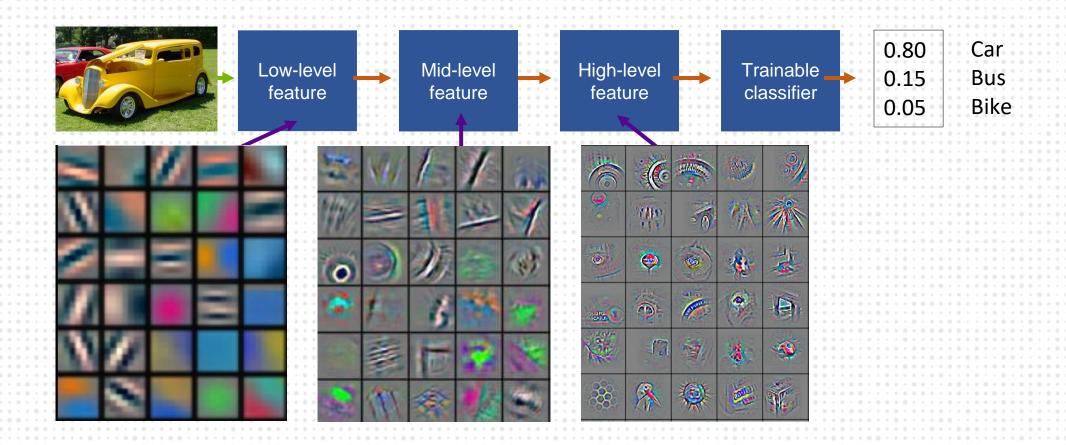
Training Data Sample size, N= 6, No. of features, F = 4)				
F1	F2	F3	F4	Y
2.1	0	400	-9	А
3.0	1	890	-42	В
2.2	1	929	0	В
4.0	0	324	-23	А
3.5	1	333	-15	А
6.0	0	215	-9	А



Key parameters of Random Forest Model are: (a) Number of trees, (b) Maximum depth of the trees (c) Size of the random subset of features In this example, No. of trees = 4, Depth = 2, and Feature subset size, m = 2 (no. of features/2)

Source: AL.com

CNN for Classification



Translational Al Center

Interactive Session

- Download notebooks from Google Colab
- Find libraries along with versions needed
- Transfer from the local drive to Nova
- Create a Conda environment and install the libraries
- Run the notebooks and check time
- Convert to Python script and run parallel (random forest)
- Run on CPU vs GPU (CNN)



Workshop materials

https://github.com/znjubery/hpc-ai-intro_2025