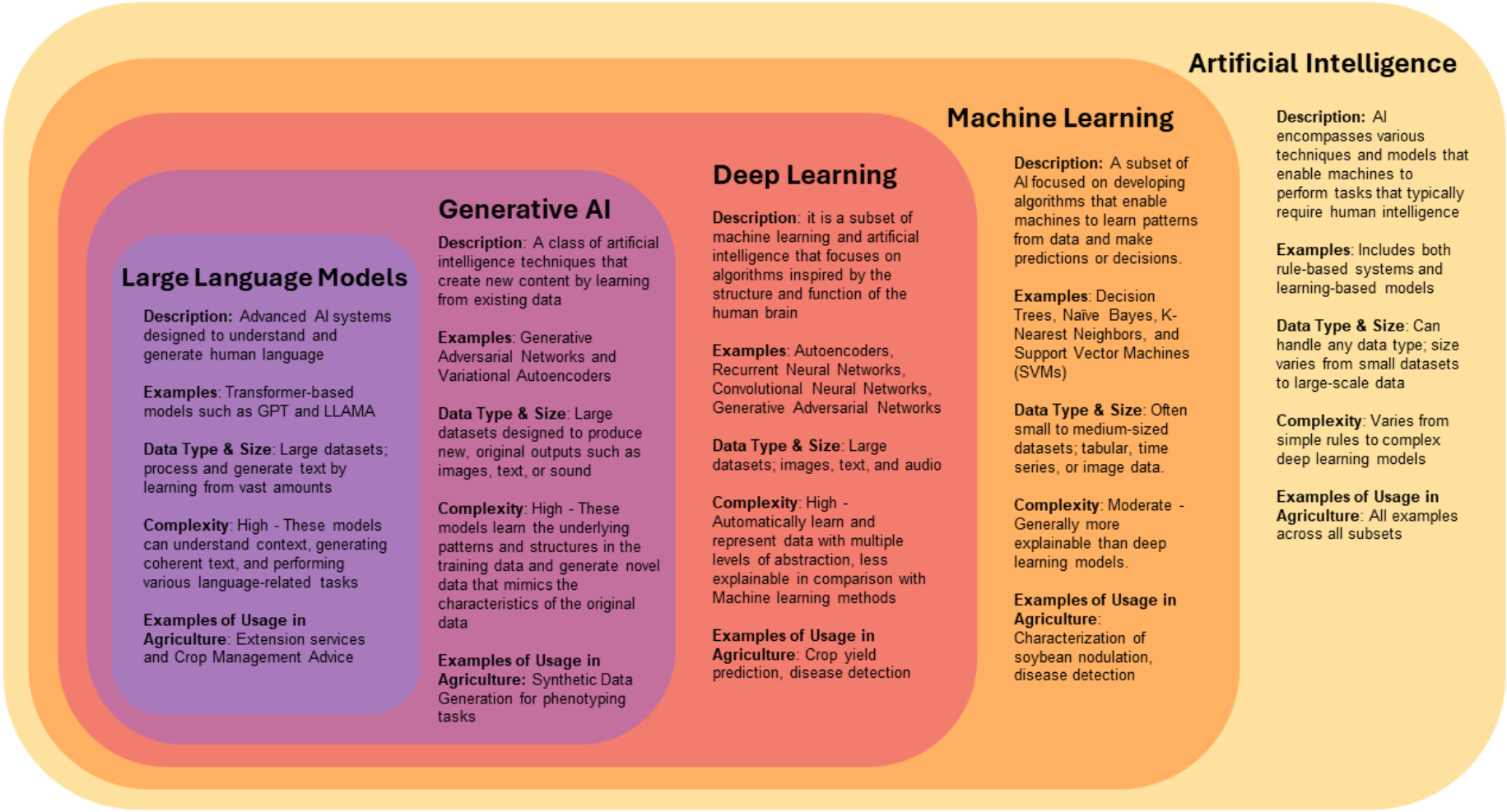
IOWA STATE UNIVERSITY

Translational AI Center

Intro to ML/AI with HPC

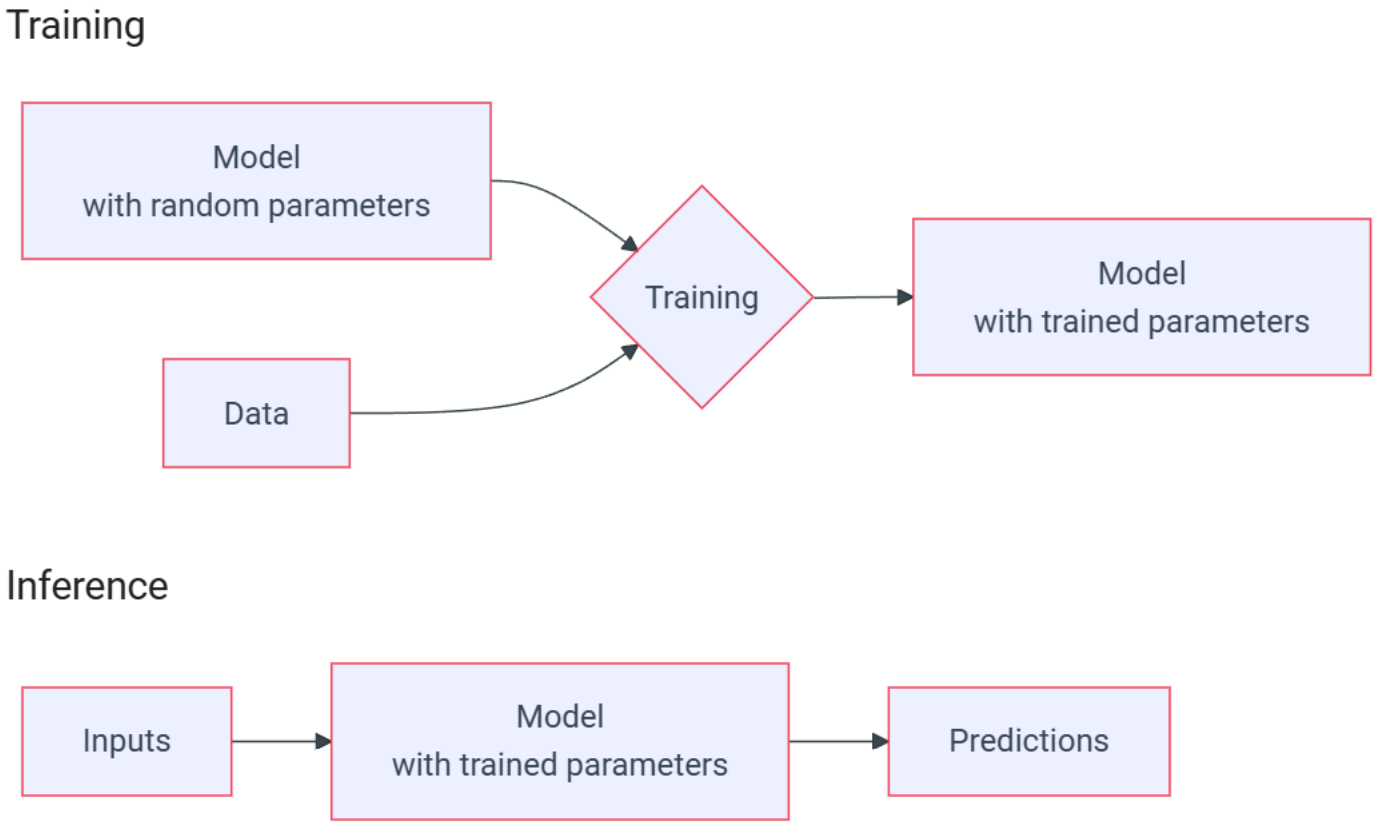
AI

"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



Singh et al. 2024, Use of Artificial Intelligence in Soybean Breeding and Production

Machine Learning Workflow



https://hpcdocs.hpc.arizona.edu/

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

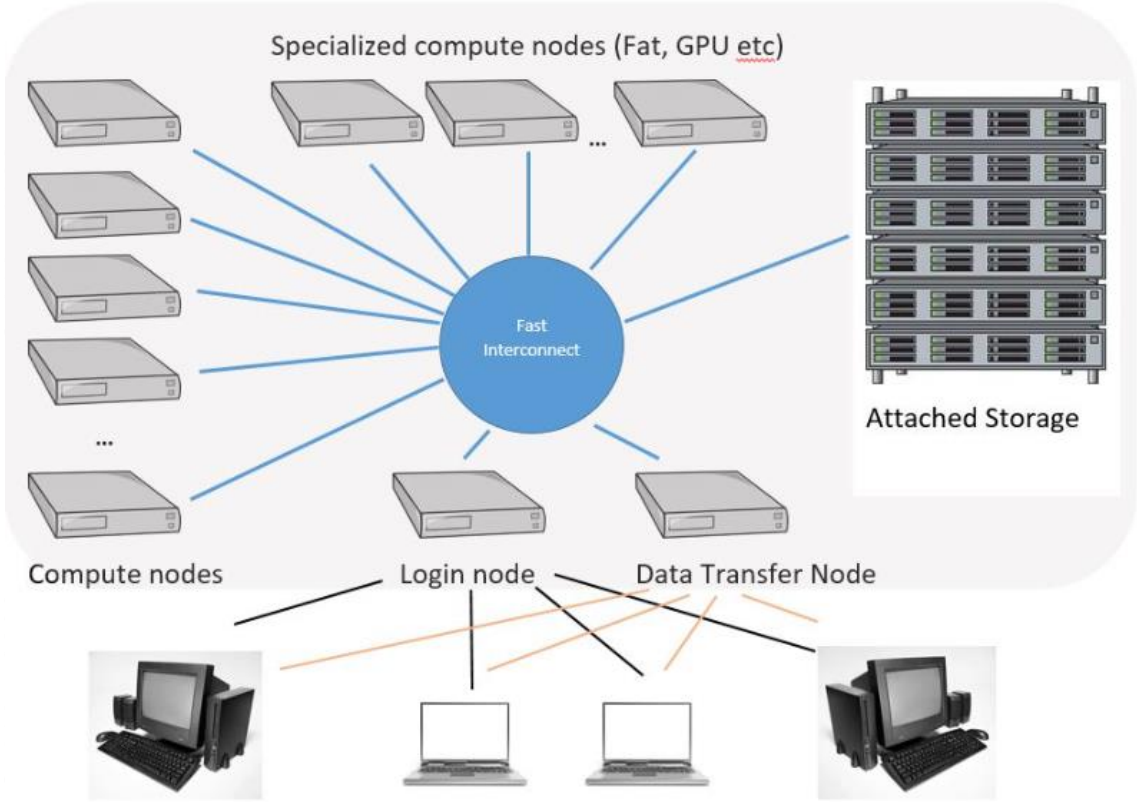
Cluster data into different groups

Write poetry, create art

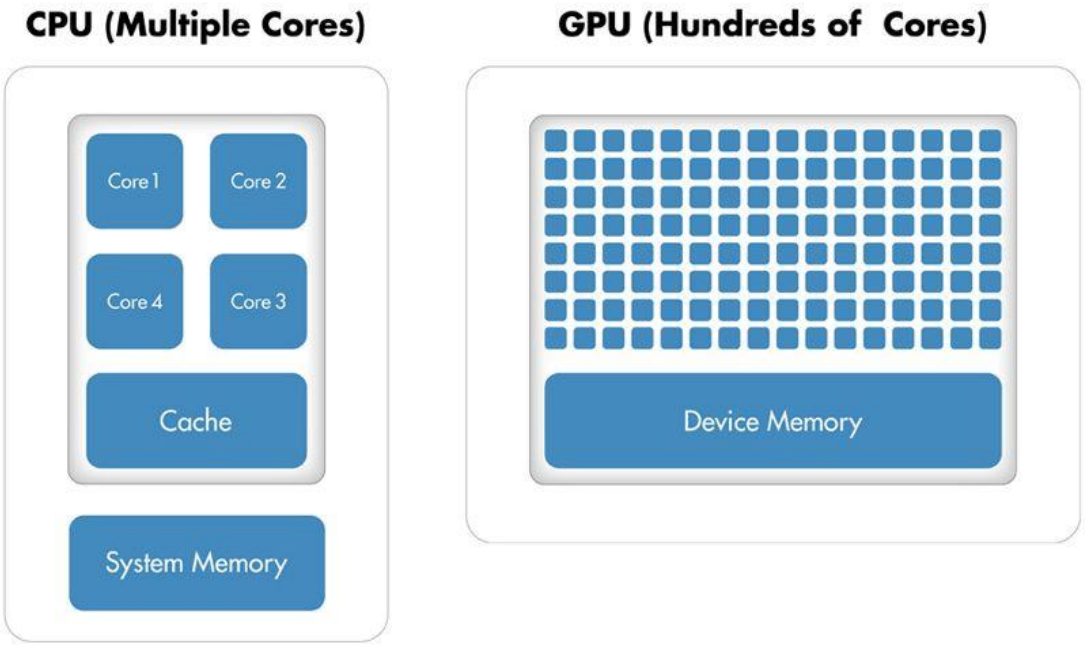
Things to keep in mind

- 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



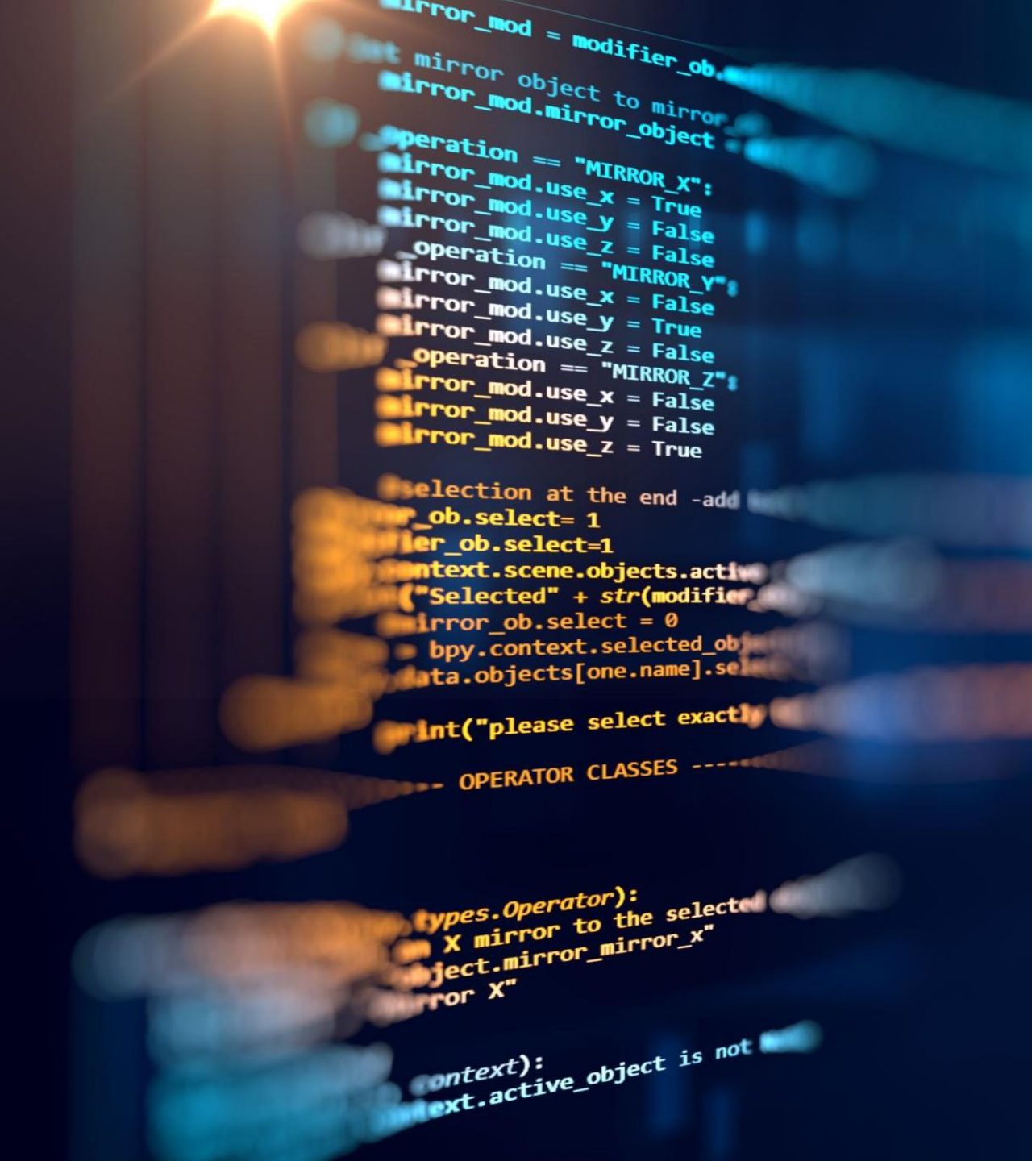
https://www.hpc.iastate.edu/



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



Best Practices for Using HPC for Machine Learning

Choose: The Right HPC System

Optimize: Code for HPC

Use: Parallel Processing

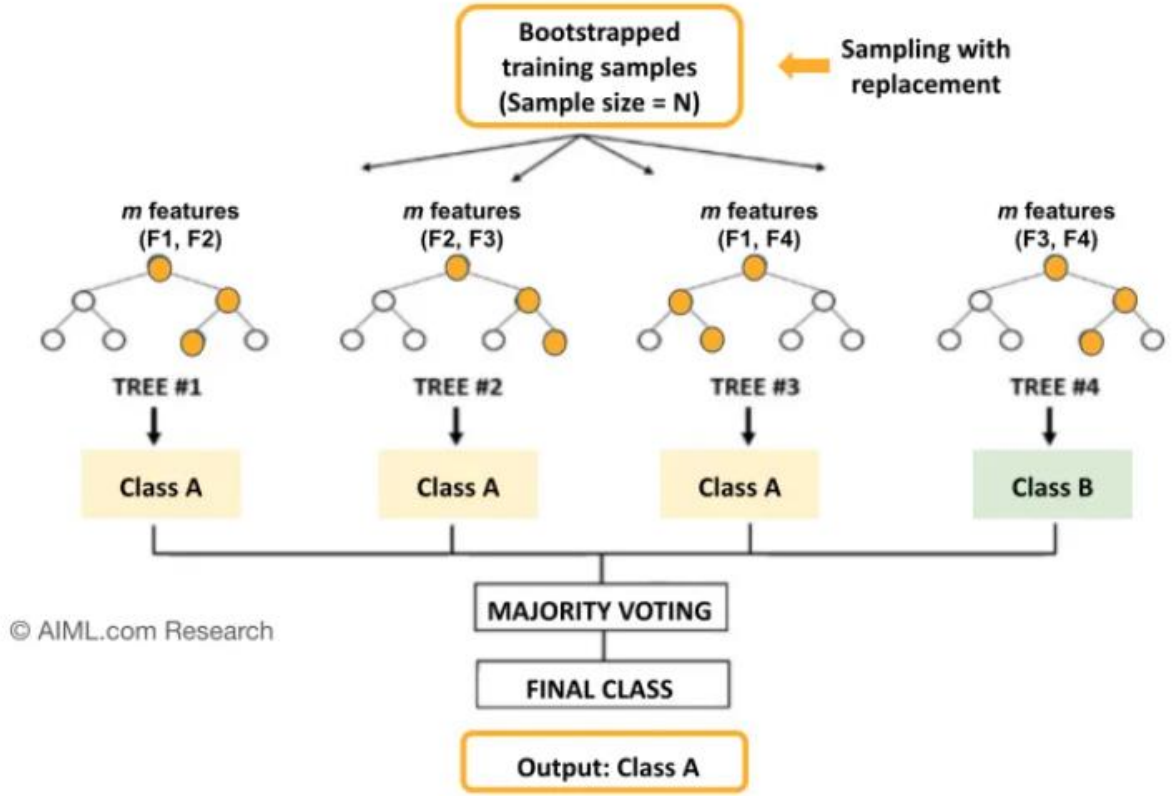
Monitor: Performance

Use: Machine Learning Optimized Libraries

Random Forest for Classification

Training Data
(Sample size, N= 6, No. of features, F = 4)

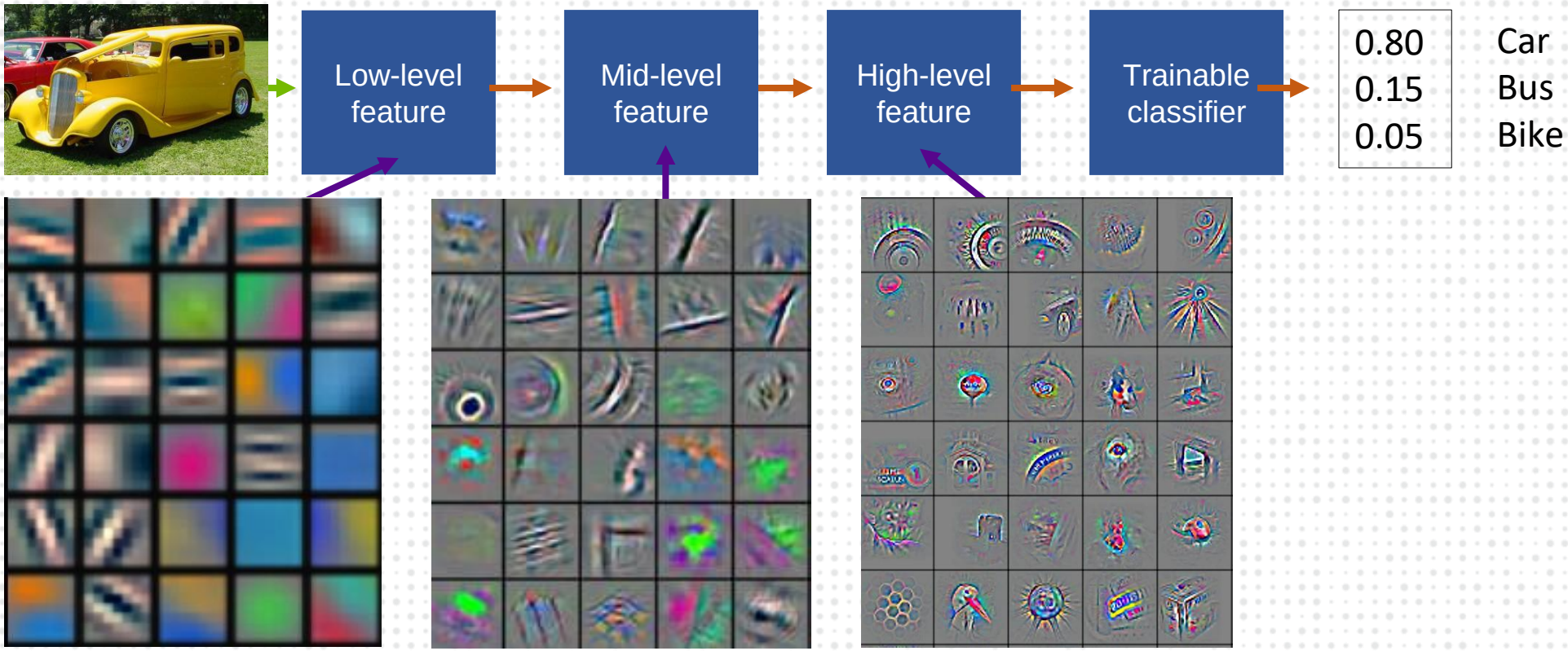
F1	F2	F3	F4	Y
2.1	0	400	-9	A
3.0	1	890	-42	B
2.2	1	929	0	B
4.0	0	324	-23	A
3.5	1	333	-15	A
6.0	0	215	-9	A



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



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

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