Advanced AI Methods

Bethany Lusch
Assistant Computer Scientist
Argonne Leadership Computing Facility
Argonne National Laboratory
blusch@anl.gov

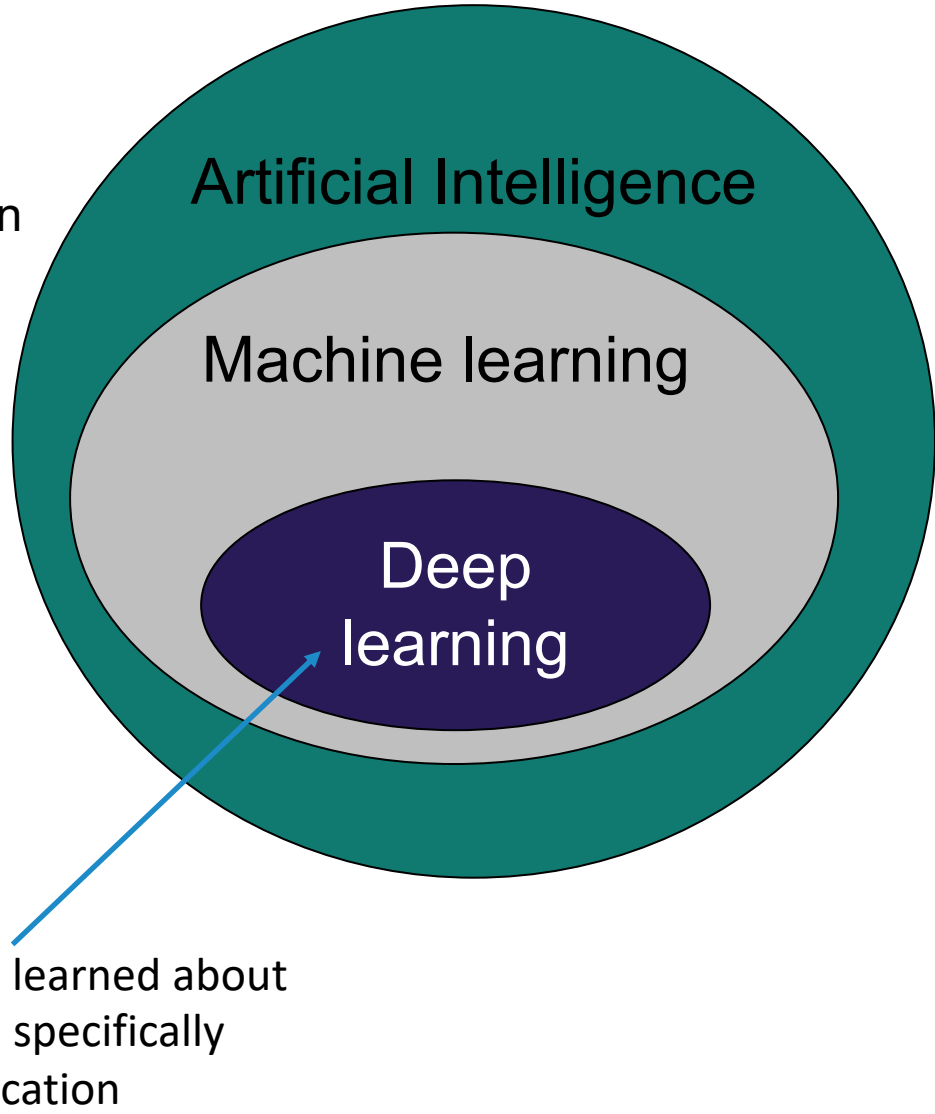
Intro to AI-driven Science on Supercomputers
November 8, 2022

Recapping Methods

What we've used so far:

- Week 1: linear regression and k-means clustering
- Weeks 2-7: neural networks (deep learning) for image classification
  - Week 2: build out mathematics by hand in Python
  - Week 3: introduce TensorFlow and convolutional layers (CNNs)
  - Week 4: more modern/advanced CNN
  - Week 5: data pipelines instead of all data in memory
  - Week 6: parallel training
  - Week 7: large-scale training (parallel & data pipeline)

Today: Gaussian processes & an intro to other AI methods



Homework This Week

Submit a paragraph about:

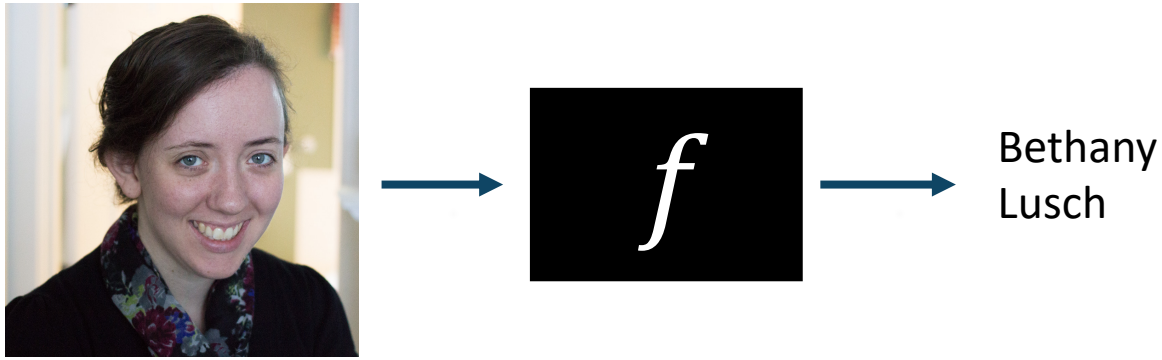
- How could you use AI for a problem that interests you?
- What is the task?
- What kind of data would you use?
- What kind of method or model might be appropriate?
- What kind of metric would you use to measure success?
- Feel free to consult the Internet for ideas

Categories of Tasks/Approaches

Classification

Have a category label for each data point, learn to categorize

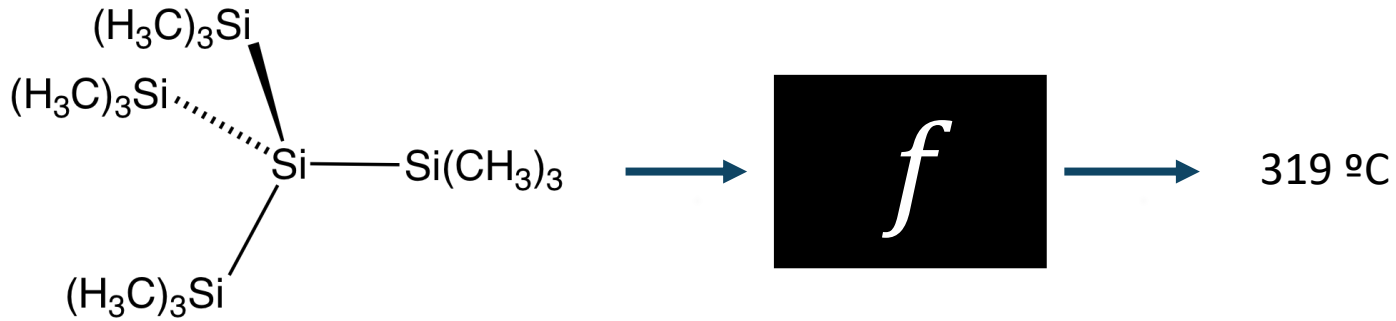
- Learn how to tag Facebook photos with the right name (after we tag many other photos of our friends)
- Learn how to label x-ray images with a diagnosis (after seeing many images labeled by experts)



Regression

Have a numeric label for each data point, learn to predict number

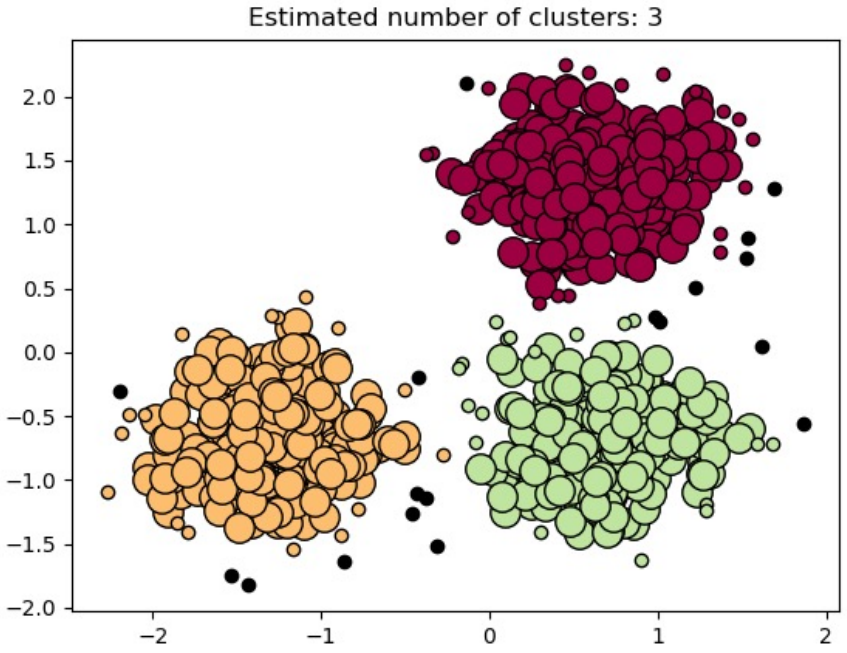
- Learn how to predict stock prices (after seeing historical stock data)
- Learn how to predict the melting point of a molecule (after seeing lots of experimental data)



Clustering

Have an unlabeled dataset, find groups of similar data points

- Find subtypes of breast cancer (after seeing data from a bunch of patients)
- Find communities in a social network (after seeing Twitter data)



Reinforcement Learning

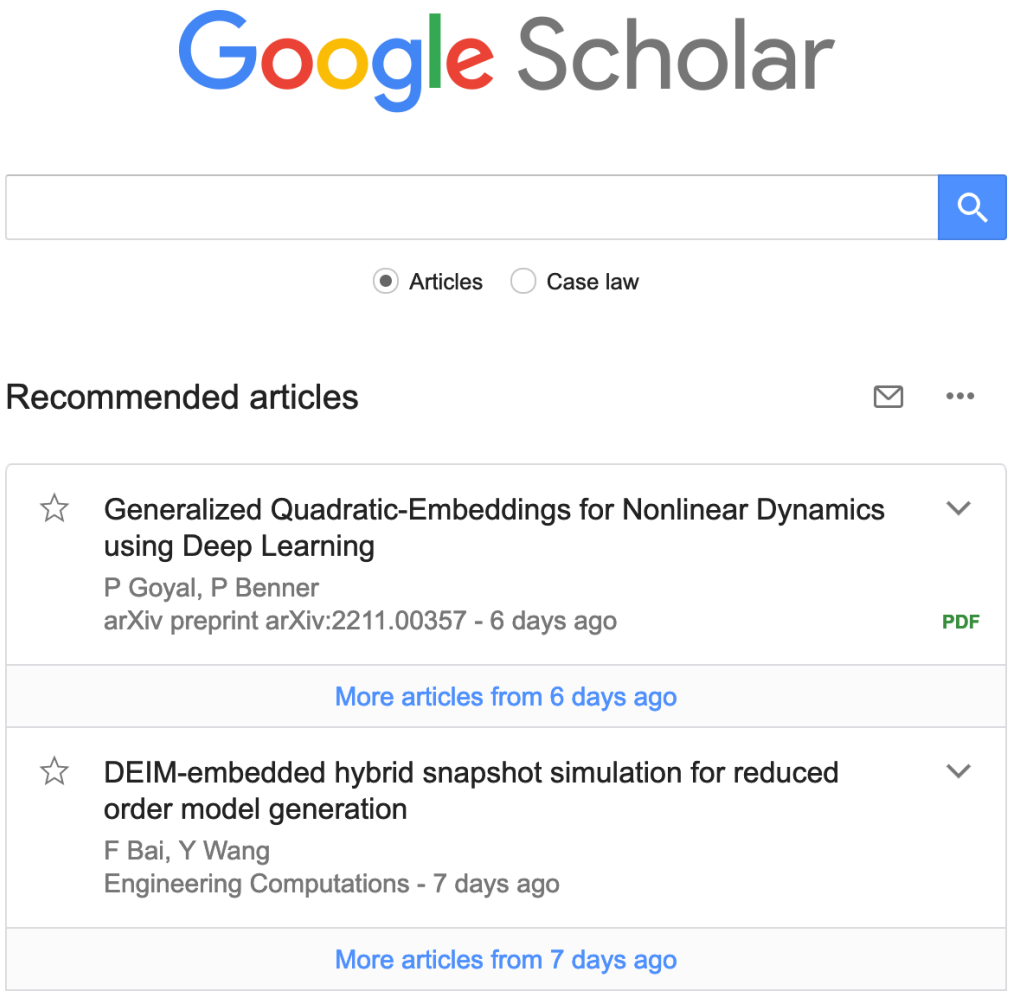
An agent explores an environment and learns how to get rewarded

- Learn to play Frogger by playing the game and receiving feedback (score)
- Learn to suggest useful chemical reactions

Recommendation Systems

Generate personalized suggestions

- Recommend a movie to watch on Netflix
- Recommend a scientific paper to read



Generative Modeling

Create new examples from a probability distribution

- Generate an image based on a text description
- Super-resolution for physics simulation

“a painting of a fix sitting in a field at sunrise in the style of Claude Monet”
DALL-E 2 (openai.com)

Supervision

- Supervised learning: have labeled data (input and output pairs) Examples: classification & regression
- Unsupervised learning: have unlabeled data and want to find patterns, structure, etc. Example: clustering
- Semi-supervised learning: mix of labeled & unlabeled data

Examples of Other AI Goals

- Intelligently **searching** through many solutions, such as planning how to a robot should complete a task
- Automated **reasoning**, such as theorem proving
- Natural **language** processing, such as writing an essay based on a prompt
- **Perception**, such as speech recognition and computer vision
- **Dimensionality reduction,** such as finding the primary modes within a fluids simulation
- **Anomaly detection**, such as flagging suspicious behavior in computer networks

Methods

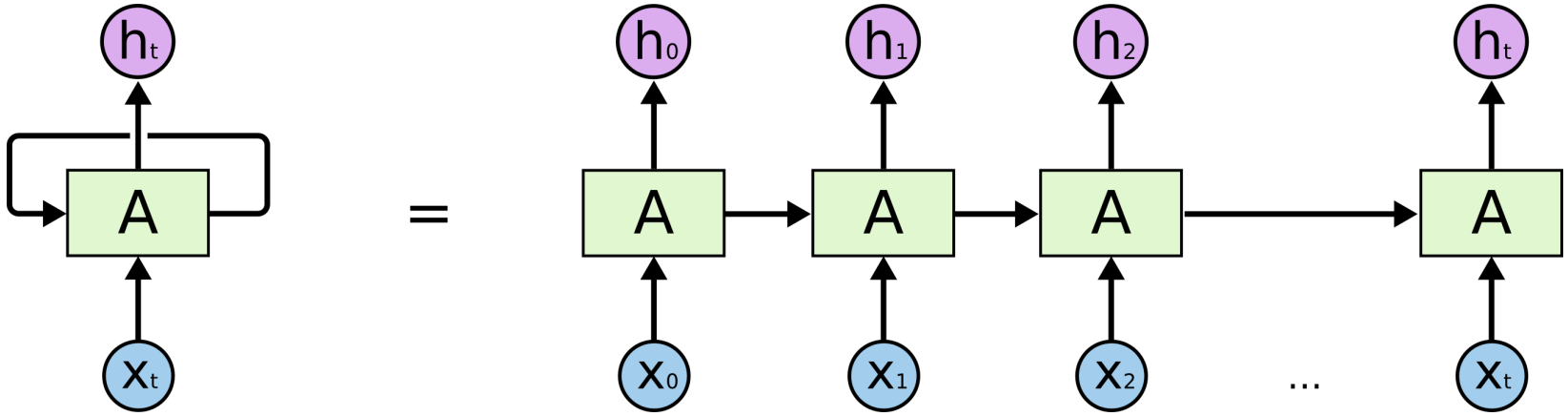
Variants of Neural Networks

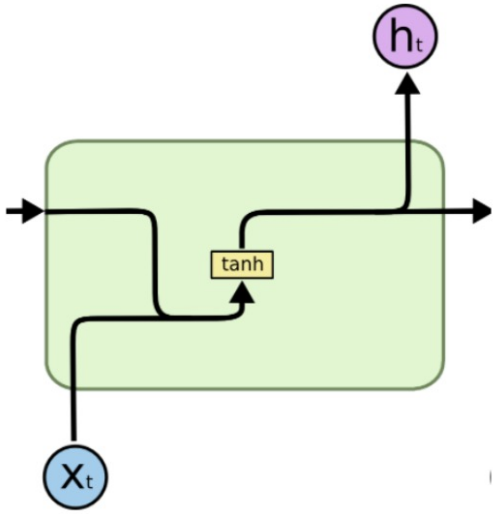
- For time series prediction
  - Recurrent neural network layers, such as LSTM and attention
  - Popular now: “Transformer” models
- To incorporate known physics, such as
  - Invariances
  - Conservation laws
  - Known PDEs (dominant method: “PINNs”)
- For generative modeling
  - Generative Adversarial Networks (GANs)
  - Diffusion models
- For language:
  - Often also recurrent, operating on a sequence
  - “Large language models” using transformers
- For dimensionality reduction:
  - Autoencoders have a bottleneck layer and then reconstruct the input

See: bonusMaterial/04_images_time_series

See: bonusMaterial/07_physics-inspiredAI

Simple Recurrent Neural Network (RNN)



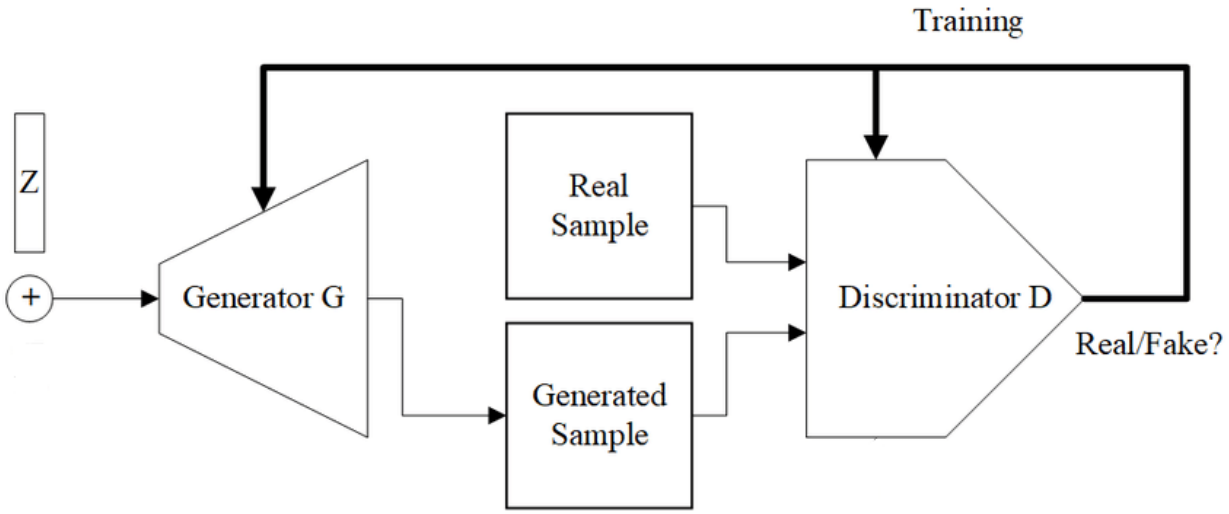


Modern RNNs have improvements on this core idea

Images from Chirstopher Olah's 2015 blog post, [Understanding LSTM Networks]

Generative Adversarial Networks

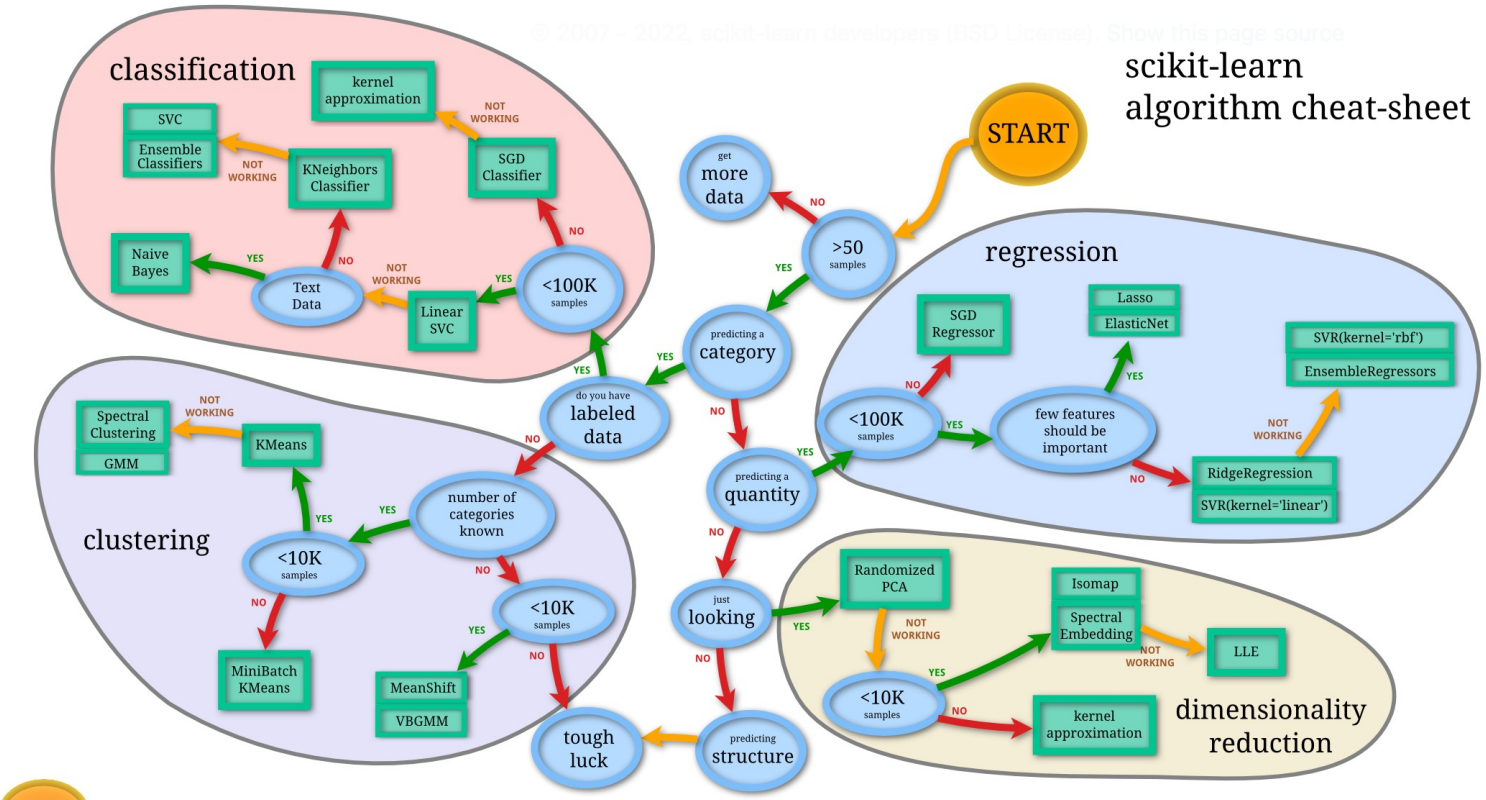
- A generative model takes random input and outputs a realistic example (such as a handwritten digit)
- A discriminator model classifies its input as real or fake
- The joint training improves both models



Source: Wikipedia האדם-החושב

“Classical” Machine Learning

- Non-deep machine learning
- Many methods are in the Python package scikit-learn

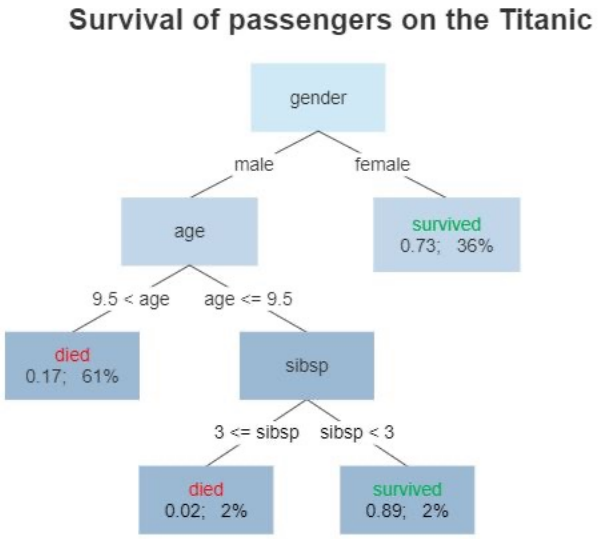


Source: scikit-learn.org

Decision Trees

- Travel down the tree, which splits based on one feature at a time
- Often helpful for features that are not continuous
- Ensembles of decision trees with some randomness:
  - Random Forests
  - XGBoost (gradient boosting)

Like having a bunch of imperfect experts voting



Source: Gilgoldm on Wikipedia

Genetic Algorithms

- Inspired by natural selection
- Start with candidate solutions and see how well they do
- Next generation: mutate the most “fit” solutions
- Continue for certain number of generations or until sufficiently “fit”

Example:

- Initial guesses for good hyperparameters
- Keep tweaking
- Want good accuracy with neural network

Augmenting Techniques

Automated Machine Learning

- Hyperparameter optimization: search for good hyperparameters such as learning rate, batch size, etc.
- Neural Architecture Search: search for a good architecture, even a very unusual one
- Even data cleaning, choose a good ML method, etc.

Improving Datasets

- Active learning: choose new training data to label that will improve the ML model
- Data augmentation: make dataset larger by, for example, including all rotations of existing images
- Subset selection: find a subset of your data examples that is sufficient for good accuracy
- Feature selection: reduce the dimensionality of your examples by removing useless features

Conclusions

- AI is a fast-moving field with broad applications
- Supercomputing can help handle large datasets and models
- There is a lot left to learn and explore!