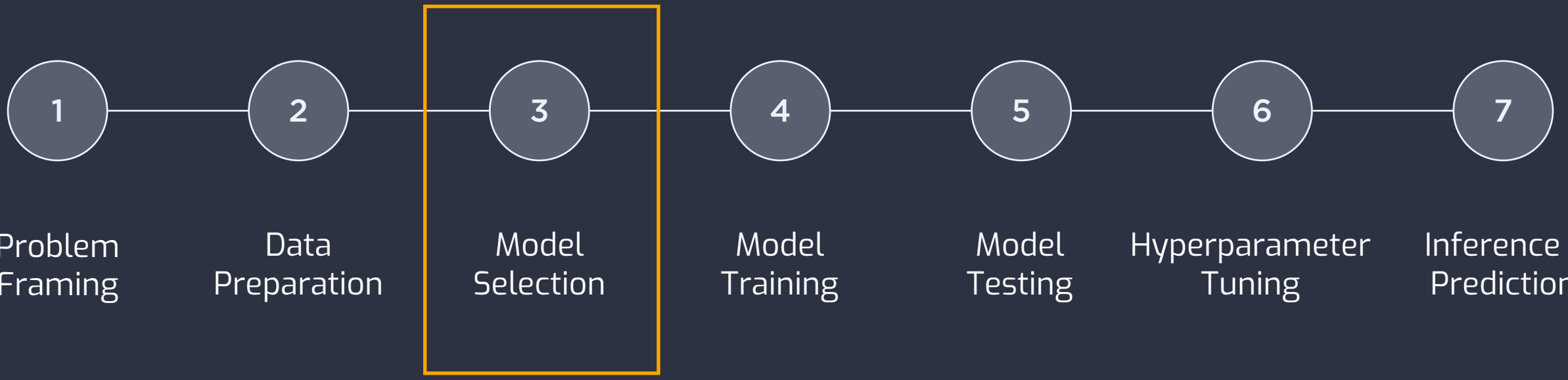
COMP2261 ARTIFICIAL INTELLIGENCE / MACHINE LEARNING

Overview of Machine Learning Algorithms

Dr SHI Lei

Durham University

learning lab


1
2
3
4
5
6
7
Problem Framing
Data Preparation
Model Selection
Model Training
Model Testing
Hyperparameter Tuning
Inference / Prediction

Learning Objectives

- Understand machine learning algorithms vs models
- Have an overview of machine learning algorithms

Grouping Machine Learning Algorithms

Based on what kind of inference we want to see in our data...

Classification	Regression	Clustering	Dimensionality Reduction

Useful for deciding which models / algorithms to try for a particular machine learning problem

In a particular problem domain,

- What do we want to do with our data set?
- What do we need to do in order to pre-process our data for training / fitting the most appropriate model?

Grouping Machine Learning Algorithms

Based on the function similarity...

Bayesian Algorithms

Regression Algorithms

Association Rule Learning Algorithms

Regularisation Algorithms

Dimensionality Reduction Algorithms

Instance-based Algorithms

Ensemble Algorithms

Clustering Algorithms

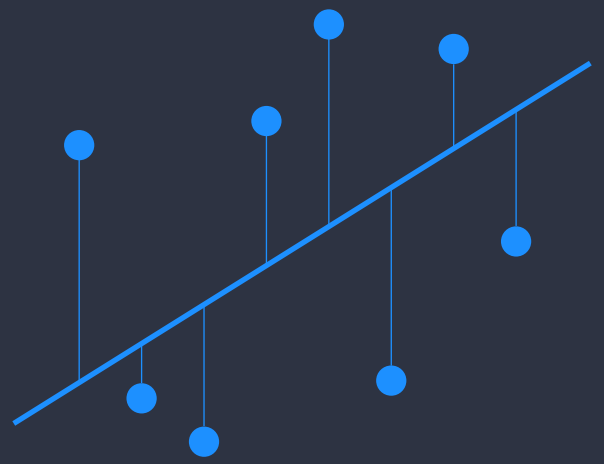
Artificial Neural Network Algorithms

Tree-based Algorithms

Deep Learning Algorithms

Regression Algorithms

- To find an approximation function mapping input variable X to continuous output variable y.
- Output variable y can be real values i.e. integers or floating-point values.
- To predict quantities, height, weight, sizes, etc.
- Learning in iterations, using a measure of error in the predictions made by the model.



The most popular regression algorithms include:

- Linear Regression
- Logistic Regression
- Stepwise Regression
- Ordinary Least Squares Regression (OLSR)
- Multivariate Adaptive Regression Splines (MARS)
- Locally Estimated Scatterplot Smoothing (LOESS)

Regularisation Algorithms

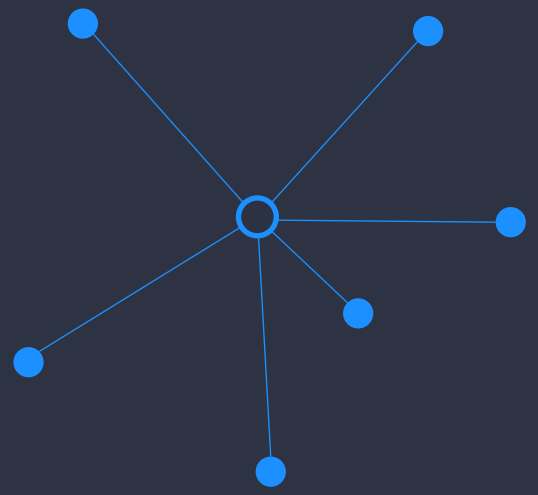
- Regularisation is a form of regression.
- Regularising / shrinking the coefficient estimates towards zero.
- To penalise models based on their complexity and flexibility.
- Favouriting simpler models to prevent the risk of overfitting.

The most popular regularisation algorithms include:

- Ridge Regression
- Elastic Net
- Least-Angle Regression (LARS)
- Least Absolute Shrinkage & Selection Operator (LASSO)

Instance-based Algorithms

- To compare new problem instances with instances seen in training, rather than performing explicit generalisation as what model-based algorithms do.
- Storing all instances in training process, and then predictions are made by comparing the new instance to the old ones using similarity measures (risk of overfitting).
- Can adapt to previously unseen data – simply store new instances or drop old instances.

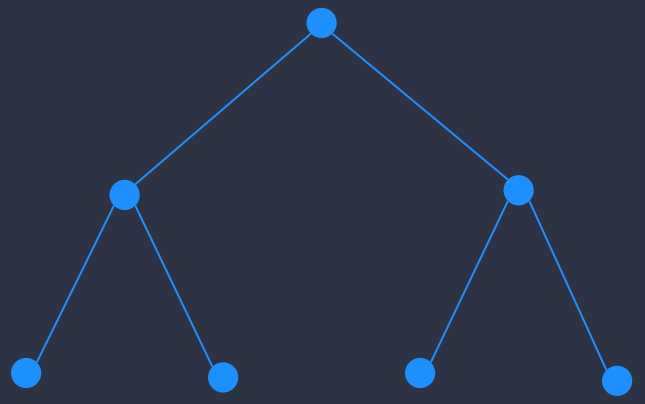


The most popular instance-based algorithms include:

- K Nearest Neighbor (KNN)
- Self-Organizing Map (SOM)
- Learning Vector Quantization (LVQ)
- Locally Weighted Learning (LWL)

Tree-based Algorithms

- To construct tree-based models of decisions made based on actual values of attributes in the data.
- The decision flow goes along a tree structure until it reaches the leaf node, i.e. the final decision for a given instance.
- Straightforward to interpret even for non-technical people.
- Large decision trees are complex, time-consuming and less accurate in predicting outcomes.

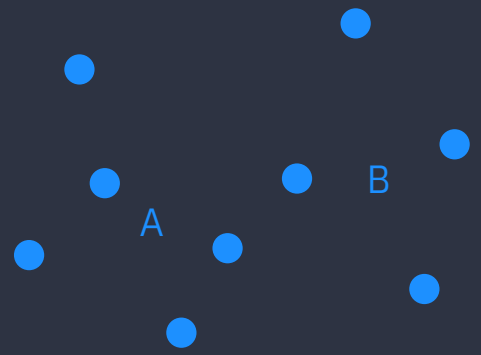


The most popular tree-based algorithms include:

- Classification and Regression Tree (CART)
- Conditional Decision Tree
- Iterative Dichotomiser 3 (ID3)
- Chi-squared Automatic Interaction Detection (CHAID)
- Decision stump

Clustering Algorithms

- To use inherent structures in the data to automatically group data for maximum commonality.
- No explicit labels for clusters/groups of instances.
- To group instances based on their centroid, density, distribution, or hierarchy
- Unsupervised learning to find natural groups in the feature space of input data.

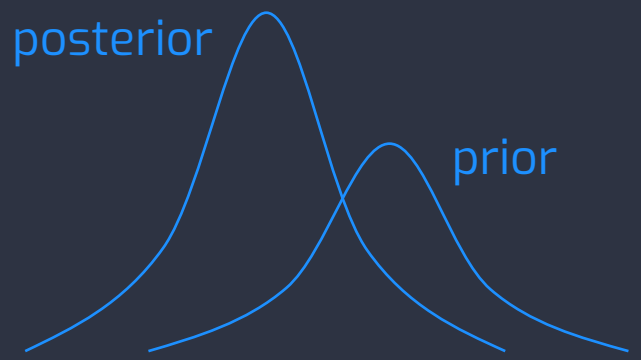


The most popular clustering algorithms include:

- K-Means
- K-Medians
- Mean-shift
- Hierarchical Clustering
- Fuzzy Clustering
- Expectation Maximisation (EM)

Bayesian Algorithms

- To create models that explicitly apply Bayes' Theorem.

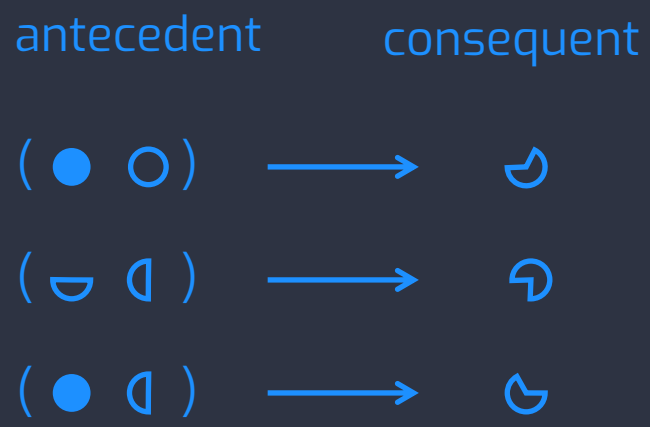


The most popular Bayesian algorithms include:

- Naïve Bayes
- Gaussian Naïve Bayes
- Multinominal Naïve Bayes
- Averaged One-Dependence Estimators (AODE)
- Bayesian Belief Network (BBN)
- Bayesian Network (BN)

Association Rule Learning Algorithms

- A rule-based to discover interesting relations between variables in large databases.
- Seeking strong rules discovered in databases using some measures of interestingness.

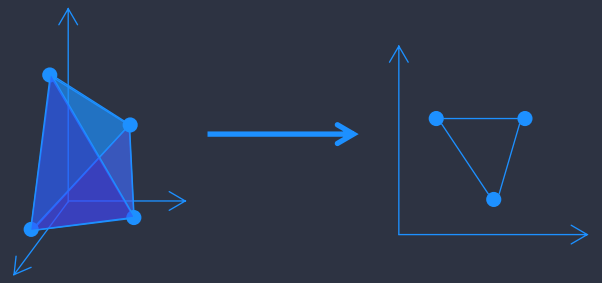


The most popular ARL algorithms include:

- Apriori algorithm
- Eclat algorithm
- FP-growth algorithm

Dimensionality Reduction Algorithms

- To find the inherent structure in data in order to reduce the number of random variables.
- Feature selection, to find the subset of original set of variables/features to get a smaller subset for a particular problem.
- Feature extraction, to reduce data in order to have a space with lesser number of dimensions.

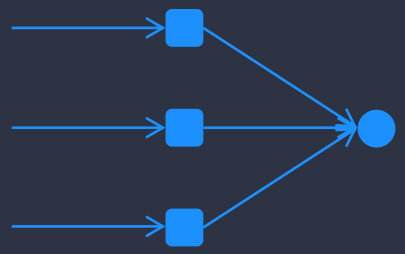


The most popular DR algorithms include:

- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Generalized Discriminant Analysis (GDA)
- Quadratic Discriminant Analysis (QDA)
- Principal Component Regression (PCR)
- Partial Least Squares Regression (PLSR)
- Multidimensional Scaling (MDS)

Ensemble Algorithms

- Combining several machine learning models into one optimal predictive model.
- To decrease variance (bagging) and bias (boosting), or to improve predictions (stacking).

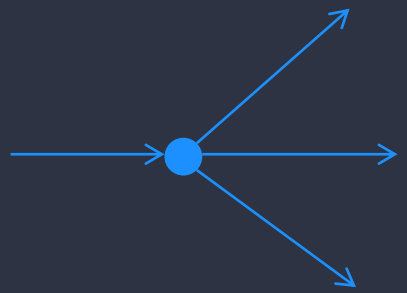


The most popular Ensemble algorithms include:

- AdaBoost
- Boosting
- Bootstrap aggregating (Bagging)
- Gradient boosted decision tree (GBDT)
- Gradient boosting machine (GBM)
- Random Forest
- Stacked Generalization (blending)

Artificial Neural Network Algorithms

- Inspired by structure & function of biological neural networks that constitute animal brains.
- A neural network is based on a collection of connected units or nodes called artificial neurons, which model the neurons in a biological brain.
- Each neuron can transmit a signal to other neurons.

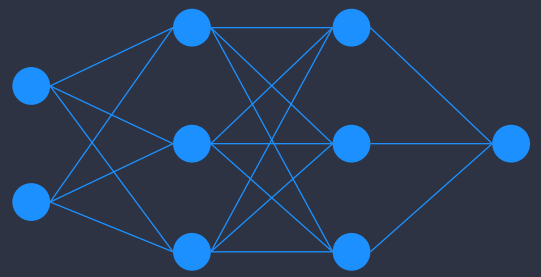


The most popular ANN algorithms include:

- Perceptron
- Multilayer Perceptron (MLP)
- Radial Basis Function Network (RBFN)
- Feedforward Algorithm
- Back Propagation
- Hopfield Network

Deep Learning Algorithms

- Extension of Artificial Neural Networks.
- To build much larger and much more complex neural networks.
- Good at dealing with unstructured dataset such as texts, images, audios and videos.

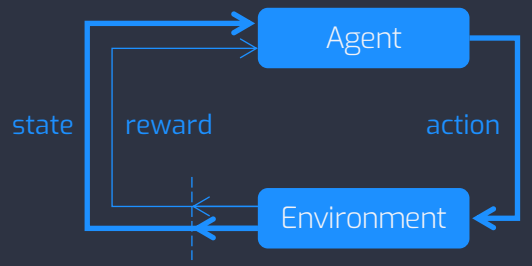


The most popular DL algorithms include:

- Convolutional Neural Networks (CNNs)
- Recurrent Neural Networks (RNNs)
- Long Short-Term Memory Networks (LSTMs)
- Deep Boltzmann Machine (DBM)
- Deep Belief Networks (DBN)
- Stacked Auto-Encoders

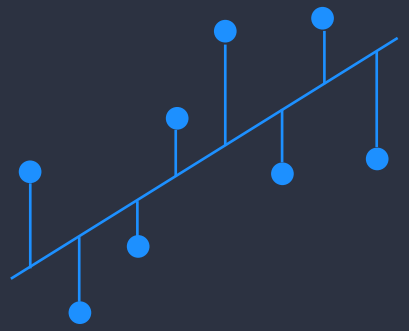
Reinforcement Learning Algorithms

- Concerned with how AI agents take actions in environment to maximise cumulative rewards.
- Agents receive rewards or punishment depending upon how they take actions in certain states.
- An environment is typically stated in the form of a Markov decision process (MDP)
- Two types of RL: model-based and model-free.



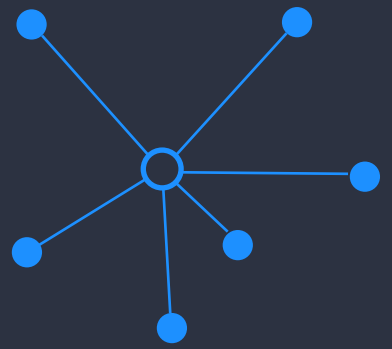
The most popular RL algorithms include:

- Model-based
  - Learn the Model: World Models, I2A, MBMF, MBVE
  - Given the Model: AlphaZero
- Model-free
  - Policy Optimisation: Policy Gradient, A2C/A3C, PPO
  - Stacked Auto-Encoders: DQN, C51, OR-DQN, HER

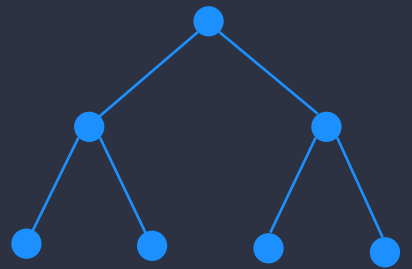


Regression

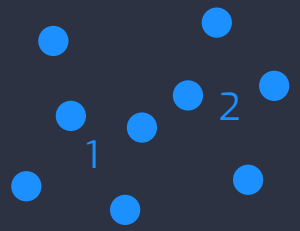
Regularisation



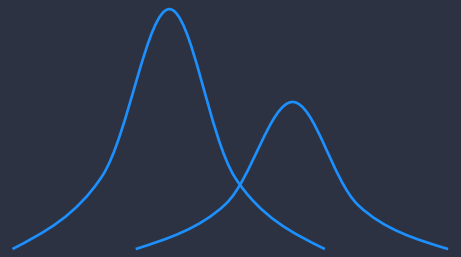
Instance-based



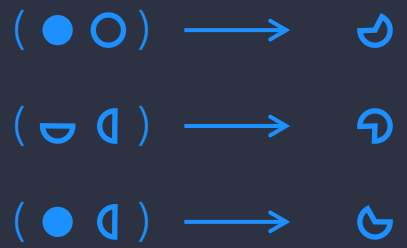
Tree-based



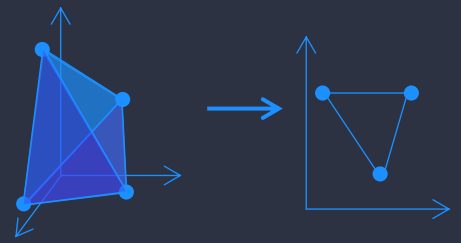
Clustering



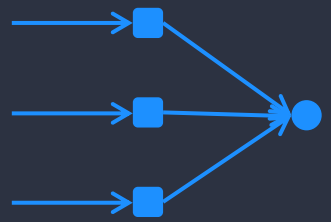
Bayesian



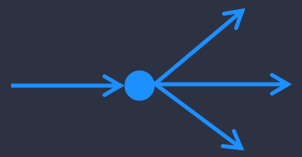
Associated Rule Learning



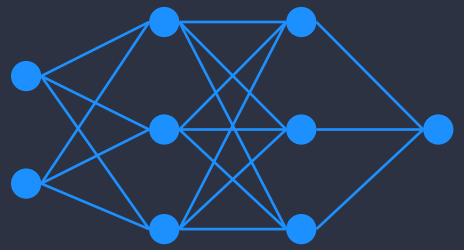
Dimensionality Reduction



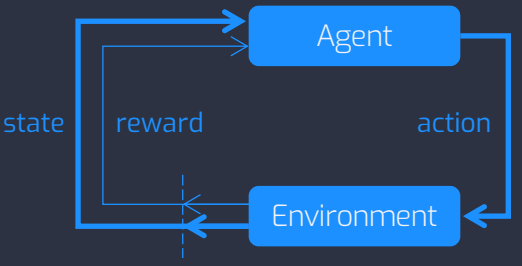
Ensemble



Neural Network



Deep Learning



Reinforcement Learning

Takeaway Points

- There are many machine learning algorithms available.
- There are many ways of grouping machine learning algorithms.
- Grouping algorithms is helpful for understanding them better and selecting them more appropriately.
- We will cover some of these algorithms in later videos, but for most of them, you will need to learn by yourself!