Carnegie Mellon University Africa
Certificate Program in AI and Machine Learning in Africa

Course AIML01: Artificial Intelligence – Past, Present, and Future

Module 2: The Nature of AI
Lecture 3: Statistical Machine Learning

Welcome to Lecture 3 of Module 2. We continue to look more deeply at the technical foundations of AI and machine learning, focusing here on machine learning, and statistical machine learning, in particular.

In this lecture, we explain the relationship between statistical machine learning and AI. We introduce three different inference strategies: analogical, domain-specific, structural. We focus mainly on different types of structural inference: regression, neural networks, Bayesian networks. We then introduce the different types of learning - supervised, unsupervised, reinforcement, self-supervised – and discuss their relationships to each other. Finally, we summarize a particular form of statistical machine learning, Bayesian learning. We will finish up by summarizing what we have covered and identifying the articles that you should read to consolidate what you have learned.
.

We have three learning objectives, so that, after studying the material covered in this lecture, you should be able to do the following.

1. Explain the connection between machine learning and AI.

2. Identify and discuss different forms of inference

3. Identify and discuss different forms of learning

Slide 1 Welcome to Lecture 3 of Module 2. We continue to look more deeply at the technical foundations of AI and machine learning, focusing here on machine learning, and statistical machine learning, in particular.

Slide 2 Over the past 20 years, machine learning, which has much in common with the connectionist approach in AI, has advanced significantly.

It is based primarily on various methods for statistical inference.

These methods use large data sets for training the system, to estimate - or learn – the parameters of a model that can then either classify new data or make predictions.

Slide 3 This approach developed in conjunction with AI research in computer vision and speech (or more generally, pattern recognition), in robotics (e.g., reinforcement learning) and in neural networks (multilayer perceptrons and deep neural networks).

The terms AI and machine learning are often used interchangeably, especially because of the strong emphasis on deep learning.

However, as we will see in what follows, machine learning retains a distinctive emphasis on data-driven statistical inference methods.

Slide 4 We can identify three inference strategies in statistics

Analogical inference
Domain-specific inference, and
Structural inference

Analogical inference maps a new problem to one that is known from previous experience (either well known or personal).

This mapping allows us to translate previous outcomes or previous solution to the new problem and infer solutions to it.

Slide 5 James Clerk Maxwell famously used an analogy from fluid dynamics (how fluid flows) to infer a model of electromagnetism, which revolutionized modern electrotechnology and communications.

Slide 6 One of the advantages of analogical inference is that it can support inferences for rare situations or situations for which there is limited data.

People use analogical inference all the time but computational analogical inference (or analogical reasoning, as it sometimes known) is less common.

Slide 7 Domain-specific inference uses techniques and knowledge that are tailored to a specific problem or environment, exploiting specialized algorithms, constraints, and domain knowledge.

Slide 8 Structural inference uses domain-general algorithms that exploit the **internal structure of the data**, rather than identifying the semantic, domain-specific, content of the data.

What the data means is irrelevant to the learning algorithm.

Structural inference is the basis of most machine learning frameworks, such as the well-known methods of regression, neural networks and Bayesian networks.

Slide 9 For example, detection of communities in complex networks often exploits patterns of connectivity rather than detailed knowledge of each entity, be they physicists who have coauthored papers on networks ...

Slide 10 or dolphins.

Slide 11 There are two classes of structural inference: logical methods and statistical methods

Both depends on the existence of statistical regularities in the dataset

Slide 12 Of these three approaches - analogical inference, domain-specific inference, and structural inference - the bulk of machine learning uses the structural inference approach.

For example, for linear regression, where the machine learns the relationship between a dependent variable (vertical axis), e.g., a person's height, and an independent variable (horizontal axis), e.g., a person's age.

Slide 13 Or for logistic regression, where the machine learns a model of a binary two-valued dependent variable, e.g., passing an exam and not passing an exam.

The probability of the outcome (dependent variable) is shown on the vertical axis, e.g., probability of passing an exam.

The values of the independent variable are shown on the horizontal axis, e.g., the hours spent studying for the exam.

Here the logistic curve is fitted to the data, and it characterizes the relationship between the hours spent studying and the (binary) outcome: passing or failing.

Slide 14 Structual inference is also the basis of neural networks, which we studied in previous lectures.

Slide 15 ... and Bayesian networks, also known as a Bayes networks, Bayes nets, belief networks, decision networks, and probabilistic graphical models.

They provide a powerful way of capturing the probabilistic, that is, the statistical, relationships among the entities being modelled and, thereby, allow inferences or conclusions to be drawn.

More technically, they provide a diagrammatic way of expressing how the joint distribution can be factored into the product of distributions over smaller subsets of variables, thereby simplifying an otherwise difficult modelling problem.

Bayesian networks were invented by Judea Pearl in 1985.

Slide 16 As we have seen, structural inference is driven by data.

Given this data-centric (sometimes known as “data- hungry”) approach,

the recent, easy availability of potentially unlimited data from social media and the web,

and wider access to cloud-based parallel computing systems such as GPUs (which are necessary to apply computationally-intensive statistical computations on large datasets)

can in great part explain the recent, impressive contribution of machine learning to AI, and information technology in general.

This made possible the bootstrapping of neural network technology from the shallow MLP networks only trainable with small datasets in the 80s, 90s and early 2000s, to the deep convolutional neural networks trained on huge datasets in the last 10 years.

Slide 17 Machine learning comprises a set of methods typically grouped into supervised and unsupervised techniques.

Slide 18 Supervised learning algorithms need a labelled dataset, that is, where each data point (for example, an image of a dog) is associated with a supervision signal or ground-truth (for example, the category label “dog”).

The learning algorithm has to find the parameters of the model (for example, weights of a neural network) using the error between the model’s own estimate and the supervision label.

Slide 19 Typically, the error provides an indication of how to adapt the parameters (e.g., the weights) in order to improve performance.

Slide 20 Examples of supervised learning algorithms include multilayer perceptrons, convolutional neural networks, and long short-term memory neural networks, decision trees, support vector machines, and regression.

Slide 21 Here we see an example of a support vector machine learning the linear boundary that best separates two classes.

The dividing line, or boundary, H3 separates the classes better than H2 does. H1 also separates the classes but with many misclassifications.

Slide 22 This figure illustrates how well the two classes are separated.

Slide 23 Unsupervised learning algorithms do not require a labelled dataset.

They discover the regularity in the data and their organization in separate categories.

Slide 24 Examples of unsupervised learning include the clustering algorithms such as k-means and autoencoder neural networks.

Here, the k-means algorithm randomly selects k start points in the feature space (in this case a 2-dimensional feature space spanned by hue and saturation, two attributes of color) and nominates these as the centroids of the k classes.

Slide 25 Then it assigns all points to the closest centroid

Slide 26 ... and calculates a new centroid for each cluster.

Slide 27 This process continues until there is no change in the centroid, at which point we have labelled all the points by assigning them to one of the k classes.

Slide 28 There a third type of learning: reinforcement learning.

Slide 29 And there is evidence that the brain engages in all three types:

Unsupervised learning in the cerebral cortex (which, loosely speaking, makes sense of all the sensory information we receive and theassociated motor commands)

Reinforcement learning in the basal ganglia (which is responsible for selecting the actions we take)

And supervised learning in the cerebellum (which is respoinsible for motor coordination).

Slide 30 Reinforcement learning can be considered part of the supervised approach, but where the supervision to learn a policy (for example, actions that should be taken when certain sensory conditions prevail) is guided by a reward function, e.g., good vs. bad.

The cost function (or objective function) that governs the learning is based on maximizing the cumulative sum of rewards over time.

Slide 31 Yan LeCun, one of the founding fathers of deep learning, uses the metaphor of a cake to show how these methods are related:

"If intelligence is a cake, the bulk of the cake is unsupervised learning, the icing on the cake is supervised learning, and the cherry on the cake is reinforcement learning" (LeCun 2016).

Slide 32 This analogy gave rise to some debate in the machine learning community, with Pieter Abbeel remarking

"I prefer to eat a cake with a lot of cherries because I like reinforcement learning."

Slide 33 Yann LeCun has recently extended the concept of unsupervised learning, using the terms **self-supervised** and **predictive learning** to describe the situation where the data provides the supervision.

Slide 34 This term refers to the power of unsupervised methods, such as autoencoders and word embeddings, that can automatically extract partial information from noisy or incomplete input data to predict the rest of the data.

Slide 35 Yann LeCun highlights the importance of prediction in intelligence.

Slide 36 To recap:

In supervised learning, training data, in the form of required behavior, provide error signals

In reinforcement learning, training signals are scalar rewards or reinforcement signals, and the goal is to maximize the cumulative sum of rewards over time.

In unsupervised learning, there are no training signals, and the goal is to uncover statistical regularities in the data.

Slide 37 Bayesian learning algorithms, which we already mentioned briefly, is another important approach to machine learning.

The general Bayesian framework is based on the intuition that the beliefs after observing some data are determined by the probability (prior probability distribution) of each possible explanation, given that data.

When processing a dataset, the machine learning algorithm uses Bayes' rule to calculate the correct probability distribution over the hypotheses given that data.

Slide 38 In other words, in Bayesian learning, the machine learns the probability that the hypothesis is true, given data.

This is computed from what we know, that is, the prior probabilities that each hypothesis is true,

and the conditional probabilities (also known as likelihoods) of that data occurring for each hypothesis.

Slide 39 Put another way, Bayes' Rule allows us to compute what we want to know from what we do know.

Given large datasets, the computations required for Bayesian learning become too difficult to be done analytically, which explains the recent boost of Bayesian algorithms because of easy access to parallel computational resources (Danks 2014).

To summarize:

1. Machine learning is a sub-set of AI and is one of the main reasons for the success of AI over the past 20 years.

2. There are three main forms of inference: analogical, domain-specific, and structural.

3. Structural inference is the most common and exploits the statistical patterns in the structure of the data, rather than the semantics of each data point.

4. The three main structural inference techniques are regression, neural networks, and Bayesian networks.

5. There are two main types of learning – supervised and unsupervised – with reinforcement learning and self-supervised learning forming respective subsets.

Recommended Reading

Here is some recommended reading.

Bassereau, P., Cangelosi, A., Cejkova, J., Gershenson, C., Goldstein, R., Martins, Z., Matthews, S., Ritort, F., Seager, S., Van Tiggelen, B., Vernon, D., and Westall, F., (2024) Chapter 4 Physics for Understanding Life, **Section 4.3 Artificial Intelligence: Powering the Fourth Industrial Revolution**, EPS Grand Challenges: Physics for Society at the Horizon 2050, IOP Publishing. http://vernon.eu/publications/2024_Bassereau_et_al.pdf

Shanthamallu, U. S., Spanias, A., Tepedelenlioglu, C., and Stanley, M. (2017). "A brief survey of machine learning methods and their sensor and IoT applications," 8th International Conference on Information, Intelligence, Systems & Applications (IISA), 2017, pp. 1-8, doi: 10.1109/IISA.2017.8316459. https://sensip.engineering.asu.edu/wp-content/uploads/2020/03/1_Machine-Learning-Survey-Paper.pdf

References

Here are the references cited to support the main points in what we covered today.

Danks D. (2014). Learning. In The Cambridge Handbook of Artificial Intelligence. Frankish, Keith, editor.; Ramsey, William M., 1960 - editor. https://doi.org/10.1609/aimag.v27i4.1904.

Klix, F. (2001). Problem Solving: Deduction, Induction, and Analogical Reasoning. International Encyclopedia of Social& Behavioral Sciences. Elsevier Ltd. 12123–12130. https://www.researchgate.net/publication/288151639_Problem_Solving_Deduction_Induction_and_Analogical_Reasoning