Lecture 5

Learning Algorithms - I

Rahul Bhattacharya

Machine Learning: The New Science

Part I

What is this New Science of Machine Learning?

Putting Machine Learning in the context of Artificial Intelligence (AI)

- 💻 Introduction, Lecture 0

Part II

Foundational Concepts of Machine Learning

- 💻 Lectures 1 – 12.

Part III

Solving Real Life Problems with Machine Learning

- 💻 Lectures 13 – 25.

Part IV

Machine Learning: The Changing Landscape of Knowledge

(Near and the Not So Distant Future)

- 💻 Lectures 26 – 27.

In this lecture and the three following lectures we will talk about what learning algorithms are and more importantly how can we classify them based on their capabilities and characteristics. This will not only bring clarity to the subject but will also help us navigate the vast field of machine learning more easily. Indeed, machine learning is all about learning algorithms and hence you can say that for the rest of this course we'll only be talking about learning algorithms. 😊

A **learning algorithm**, or a learner, is at the heart of machine learning. It is a very special kind of an algorithm that is used to make a machine (computer) learn from data. We can think of

a learning algorithm as an algorithm that empowers a computer to learn from the data and then generate all other algorithms that's embedded in that data. A computer learns from data via a learning algorithm and once it has learnt, acquired the relevant knowledge, it is able to generate all other algorithms that are needed to solve problems in practically every single domain of human knowledge, such as physics, engineering, finance, banking, geology, biology, physiology, etc.

We will explore the learning algorithms in greater detail from here on. As I said in the beginning, machine learning is all about learning algorithms and the good news is that they can make a computer learn anything: an image, patterns, a formula, mathematical functions and equations, human voice, just about anything from data. The bad news is that there are too many learning algorithms, literally hundreds of them and it is difficult to decide which one is good for the learning task at hand. It is therefore paramount to understand how the learning algorithms work and to how we can categorize them into various groups based on their characteristics and capabilities.

In fact, we will see that there are quite a few ways how we can categorize and group machine learning algorithms based on their characteristics and capabilities. This gives rise to a lot of jargon and a plethora of terminology. There are four broad methods by which we can classify machine learning algorithms. These methods can be briefly summarized as:

- Classification based on the nature of the algorithm, i.e. whether the algorithm is **linear** or **non-linear** in nature.
- Classification based on the form of the mapping function, i.e. whether the algorithm is **parametric** or **non-parametric**.
- Classification based on the learning paradigm, i.e. whether the type of learning is **Supervised**, **Unsupervised** or **Reinforcement** Learning.
- Classification based on the schools of thought (or, the "tribes") in machine learning, **Symbolists**, **Connectionists**, **Evolutionaires**, **Bayesians** and **Analogizers.**

The final classification of machine learning algorithms, which is perhaps the most intuitive, is mostly due to Pedro Domingos, the celebrated professor of Computer Science at the University of Washington, who talks about the five different "tribes" of machine learning.

Let's consider these one by one in this and the following three lectures. Once we are done with them, you'd get a good overview, kind of top down view, of what machine learning is about and would be ready to start off looking at various different algorithms in greater detail.

Linear and Non-Linear Algorithms

The learning algorithms can be classified as linear and non-linear algorithms based on their level of complexity. Linear algorithms drive the Linear Learning Machines. Remember, a computer within a machine learning framework is no longer just a machine. It is a learning machine, or a learner, driven by learning algorithms. For example, a Linear Threshold Unit (LTU) is a kind of learning machine that makes use of certain kinds of linear algorithms. In a linear learning machine hypothesis are formed as linear combination of input variables. (The term "hypothesis" is something that you need to put in the backburner for the moment, but keep in mind that the concept of hypothesis is closely related to the concept of function that we talked about in the earlier lectures. We will be talking about this term a lot in the coming lectures.) Anyway, let's outline what are linear algorithms.

A very Brief Note on the "Linear" and the "Non-linear"

All of you must be familiar with the term "linear". It means straight, like a straight line. No curves, no twists, no wiggles but a straight line. In common sense terms, you can think of linear as anything whose sum of the parts is equal to the whole. Tear a flat piece of paper, a flat sheet, from your note pad and it is linear. Crumple that sheet and it becomes non-linear. After you crumple the flat sheet of paper, no matter how hard you try to flatten it (or, smoothen it) you'd not be able to return it to its original state of flatness. In math terms, a linear variable is one which is raised to the power of 1, i.e. itself. If you write the equation $y = 3x + 2$, where x is the independent variable and y the dependent variable then this is a linear equation (equation of a straight line) because the independent variable, x is raised to the power of 1. If the power of the independent variable had been greater than 1 then it would become a non-linear equation, like $y = 2x^2 + 4$ or $y = x^3 + 3x + 1$. Non-linearity also arises when you have exponential terms (variables) such as, e^x or sinusoidal terms such as $\sin x$ in an equation.

Linear Algorithms

Let's us now very briefly survey the various different kinds of linear (learning) algorithms that are used in machine learning.

- **Linear Regression** – This is a linear algorithm.

  When we talk about linear regression, think about straight lines. Fitting a straight line to a collection of points on an X-Y plot (graph) is the problem of linear regression. The problem of linear regression consists of finding a linear function of the form: $y = ax + b$, where a and b are two constants whose value we need to find in order to fit the straight line. At the highest level in machine learning sits a function which needs to be determined and you can see that this straight-line equation can also be expressed as a function of the form $y = f(x) = ax + b$. In fact, instead of calling it a function, we can also call it a "hypothesis" but more of that later. But remember from this function of a straight line, we can create other functions as well, such as what is known as a "cost function", which will assume great importance when we start to go into details of learning algorithms.

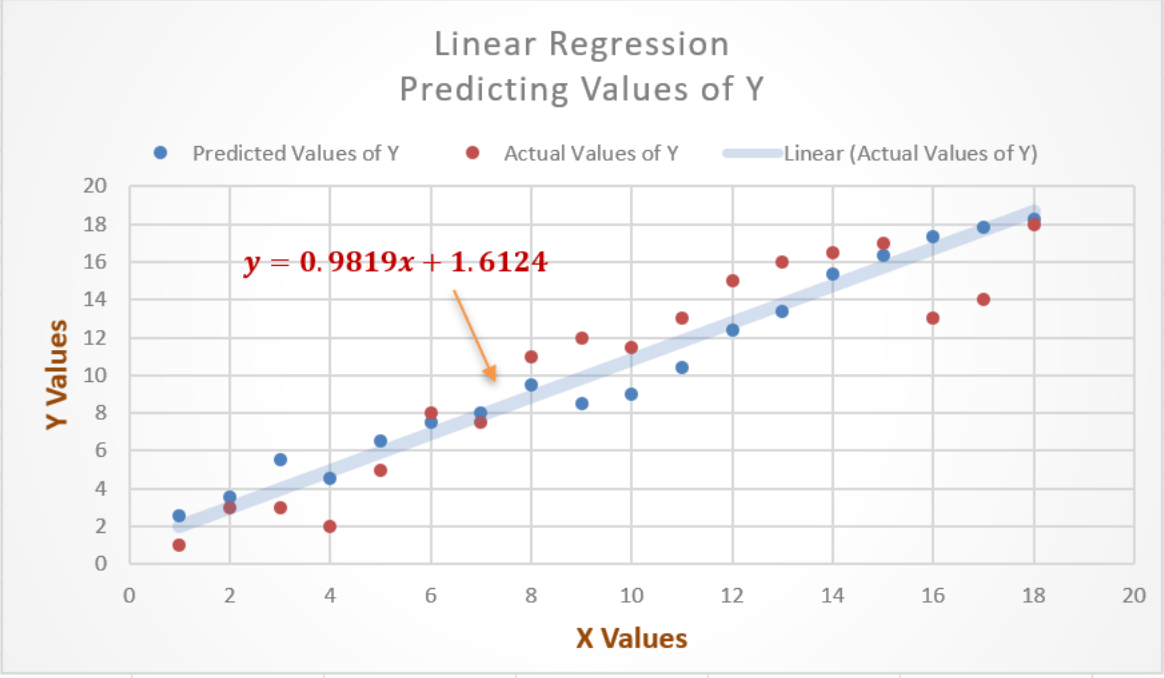

Figure 5.1 Using Linear Regression for Prediction

- **Linear Threshold Unit** – This is not a learning algorithm but rather it is a kind of Linear Learning Machine that runs on linear learning algorithms.

  Linear Threshold Units (LTU) are a kind of learning machines that make use of three different kinds of linear algorithms, Perceptron, Logistic Regression and Linear Discriminant Analysis. We would be talking about LTUs in a lot of detail in the forthcoming lectures.

  - **Logistic Regression** – This is a learning algorithm.

    This name is a bit misleading because this algorithm actually performs classification – such as, classification on problems with two classes – rather than regression. We will see in later lectures as to why we call this regression. As some of you may know, logistic function is related to the concept of probability and also to what some call the "most important curve in the world", the S-curve (also, called the logistic or the Sigmoid curve). These concepts will figure prominently in some of our later lectures. As a bit of homework, you should look up the S-curve. It is truly the "most important curve in the world" and shows up in physics, chemistry, geology, evolutionary biology, technological growth dynamics, seismology, economics, finance and almost all areas of engineering.

  - **Linear Discriminant Analysis (LDA)** – This is a learning algorithm.

    LDAs are mostly used for making classification on problems that have more than two categories. Classification is an important goal of machine learning ("high risk" and "low risk" grouping of customers in the context of loan approval by a bank is a classification problem). The objective of LDA in most machine learning applications is to project a feature space – which comprises of data in n-dimensions – onto a much smaller subspace while at the same time maintaining the class-discriminatory information. This is what is known as dimensionality reduction, i.e. we reduce the number of dimensions in which the data is represented (in an X-Y plot in a graph, there are two dimensions, one represented by x axis and the other by y axis; however, in real life problems we have problems where the data is represented on multiple dimensions, not just 2 or 3 dimensions even though we cannot graphically visualize such dimensions. It not only makes more intuitive

sense to visualize data in 2 or at most 3 dimensions but it also helps us in making calculations faster and more efficient – lower computational cost – as compared to doing the same in say, 4, 5, 10 or in even higher dimensions). Dimensionality reduction by LDA algorithm not only helps in reducing computational costs for a given classification task – after all, the learning by machines needs to be efficient – but it we can also avoid overfitting by minimizing the error in parameter estimation (some of you, especially those of you who are working may have heard of the phrase "curse of dimensionality". Once again do not worry about terms/phrases such as "overfitting" or "parameter estimation", they'll become clear to you as we move on through this course).

- **Perceptron** – This is an algorithm.

  Perceptron is one of the oldest algorithms used in machine learning used to study linear threshold units. Note that the earliest versions of perceptrons (in the 1960s) did not possess the capability to learn and hence were not strictly learning algorithms (you may want to recall our discussion on logic gates and computers how even though the operations of the logic gates – AND, OR and NOT – gave rise to some kind of internal reasoning within the computer which helped it to run fast computations and follow human instructions this still did not make the computer a learning machine as the operations of the logic gates did not give rise to learning; the earliest perceptron algorithm was also something similar to the logic gates and their combination, the only difference was that it was being presented as a likely model of neurons inside human brains rather than electronic computers.) Interestingly enough, the Perceptron was invented by a psychologist named Frank Rosenblatt in the late 1950s who was then working at Cornell University. You should bear in mind that the invention of perceptron by Rosenblatt unleashed the machine learning revolution. Even more interesting to note is that the invention of a technique known as "Backpropagation" (of which, will talk a lot when we talk about artificial neural networks) when combined with the perceptron in the mid-1980s gave it the power to solve real world problems that were governed by non-

linear models (from XOR – exclusive OR in computer science – to predicting stock markets).

- **Gradient Descent** – Strictly speaking, this is not a machine learning algorithm. It is an optimization algorithm used in mathematics

  While gradient descent algorithm is not a machine learning algorithm, it is used extensively in machine learning, primarily in conjunction with the linear learning algorithms such as linear regression and perceptrons, to solve classification and regression problems (where we need to make predictions). We need a whole new lecture, or perhaps even two, to talk about this important algorithm but suffice it to know for the moment that gradient descent algorithm that minimizes functions using an iterative method.

We have talked about a lot of important concepts in machine learning in this lecture and you need to process all this information in head. As I keep saying, do not be daunted by the jargon and terminology that you encounter. Machine learning – believe it or not – is far simpler than say, math, physics or biology. It is also much simpler than its parent, computer science.

References:

- Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev. 1958;65(6):386–408.
- Minsky, M. and S. Papert, *Perceptrons: An Introduction to Computational Geometry*, The MIT Press, 1969
- Domingos, Pedro, *The Master Algorithm*, Basic Books, 2015
- Khanna, Rahul, Mariette Awad, *Efficient Learning Machines, Theories, Concepts, and Applications for Engineers and System Designers*, Apress Open, 2015
- Abu-Mostafa, Yaser S., Malik Magdon-Ismail, Hsuan-Tien Lin, *Learning from Data, A Short Course*, AMLBooks.com, 2012.
- Raschka, Sebastian, Linear Discriminant Analysis, Bit by Bit, https://sebastianraschka.com
- Blum, Avrim, *Machine Learning Theory*, Carnegie Mellon University, 2009
- Domingos, Pedro, *The Master Algorithm*, Basic Books, 2015

Machine Learning: The New Science is an **online course** developed by Risk Latte AI to increase awareness and a very basic understanding of the field of Machine Learning, the new science that is taking our world by storm. In the next ten years, the world that we see around us would have changed completely, thanks to machine learning. In the next 20 to 30 years at most we, the human beings, would be in a totally uncharted territory, once again thanks to machine learning and the much bigger and related discipline of artificial intelligence (AI). In more ways than one, machine learning is the stepping stone to understanding artificial intelligence. Machine learning is a very important sub-field of artificial intelligence.

This online course prepares a person to start off with a more formal **Machine Learning 101** course. This course contains very easy to understand lectures and simple tutorials implemented in Microsoft Excel™ to illustrate real life problem solving using machine learning algorithms.

- This course is primarily targeted towards high school, college and university students and middle to senior level working professionals and executives around the world who have a very basic knowledge of mathematics and a very limited or even **no knowledge** of computer programming.
- For certain segments of our target audience, such as certain specific groups of high school, college and university students in India and other developing countries this online course costs only US$8.00 (USD Eight only).
- This course comes with on-demand and customized onsite, classroom lectures and workshops at a minimal cost.
- A lot of the lectures and many Excel spreadsheets containing worked examples are **FREE** and can be easily downloaded from our website. These convey the general flavor of the course.

Risk Latte AI is a unit of **Risk Latte Americas Inc.**, a Montreal, Canada incorporated company, that is in the business of developing machine learning algorithms and software for banks, financial institutions, healthcare and the education industry as well as developing gamification and social learning applications for the general public.