MINDSHARE
BRINGING LIFE TO KNOWLEDGE

1-512-256-0197
www.mindshare.com
training@mindshare.com

AI and Deep Learning Demystified

Training

Let MindShare Bring "AI and Deep Learning Demystified" To Life For You

Deep Learning Description

The major tech giants (e.g. Google, Amazon, Facebook, Microsoft, and Apple) are convinced that artificial intelligence (AI) will transform the world in short order. AI algorithms determine our internet search results. They recognize faces in our photos and identify commands spoken into our smartphones. They translate sentences from one language to another and can defeat the world champion at Go. They power self-driving cars and autonomous robots, drug discovery and genomics. In other words, the number of applications of AI technology is exploding. Investments in AI will also discover a new frontier for computing hardware, since many companies, from giant semiconductor companies to a myriad of startups, are racing to develop new AI-specific chips.

Modern AI technology is based on deep learning algorithms. These algorithms learn tasks on their own by analyzing a very large amount of data. To do so, they flow data through multiple processing layers, each of which extracts and refines information obtained from the previous layer. The "deep" in deep learning refers to the fact that these algorithms use a large number, say dozens or hundreds, of processing layers. This depth allows them to learn complex tasks.

Course Overview

Deep learning algorithms are remarkably simple to understand and easy to code. Through a sequence of hands-on programming labs and straight-to-the-point, no-nonsense slides and explanations, you will be guided toward developing a clear, solid, and intuitive understanding of deep learning algorithms and why they work so well for AI applications.

Practice: Three types of neural networks power 95% of today's deep learning commercial applications: fully connected neural networks; convolutional neural networks; and recurrent neural networks. During this training you will gain a solid understanding of each of these neural networks and their typical commercial applications. Most importantly, you will learn how to implement them from scratch with Pytorch (the deep learning library developed by Facebook AI). You will then train them on various image recognition and natural language processing tasks, and build a feel for what they can accomplish.

Theory: Deep neural networks are trained with gradient descent and backpropagation. These are simple yet very powerful concepts, and they provide the mathematical rules that govern the learning process. You will gain a solid, clear and intuitive understanding of these two fundamental concepts.

GPUs and cloud computing: As we need GPUs in order to train deep neural networks, the programming labs will all take place on a modern cloud platform. You will be given access to your own cloud virtual machine instance with a dedicated GPU to implement and train deep neural networks.

You Will Learn:

- Deep learning algorithms -- why they are at the center of the ongoing AI revolution, and what are their main commercial applications.
- How to implement deep learning algorithms with Pytorch (the deep learning library developed by Facebook's artificial intelligence research group).
- Fully connected neural networks, convolutional neural networks, and recurrent neural networks.
- The inside mechanics of deep learning algorithms and why they work so well.

- The hardware challenges for deep learning applications and current trends for addressing them.

Who Should Attend?

- Anyone that wants to take a first step toward using AI for their own applications.
- Anyone that will work with or around AI algorithms and wants a solid understanding of how these algorithms work and what type of tasks they can accomplish.
- Anyone that want to understand how AI algorithms are currently being used by the major tech giants (e.g. Google, Amazon, Facebook, Microsoft, and Apple), and how these algorithms will profoundly impact many industries in the near future.

Course Length: 3 days

Course Outline:

Day 1

Morning session: Introduction and background

- Introduction
  - Brief history of neural networks, deep learning, and AI
  - Recent breakthroughs and today's main commercial applications
  - What is the difference between deep learning and traditional machine learning?
  - Modern deep learning frameworks: TensorFlow, Pytorch, Caffe2

- Brief linear algebra review
  - Three ways to understand Matrix-Matrix multiplications
  - The outer product of two vectors

- Introduction to Pytorch
  - **Lab 1**: Manipulating tensors
  - **Lab 2**: Manipulating datasets

Afternoon session: A first look at neural networks

- What is a neural network? How to implement and train a neural network with Pytorch?
  - Linear layers, nonlinearities and the softmax function
  - **Lab 3, 4 and 5**: Constructing a neural network with Pytorch
  - **Lab 6**: Playing with a trained neural network
  - **Lab 7**: Training a neural network on a simple image recognition task

- A gentle and intuitive introduction to the equations and formulas of deep learning
  - **Lab 8**: Visualizing how the internal parameters of a neural network evolve during training
  - Formula to update the internal parameters: "outer product between the error and the data"
  - A very simple way to understand this formula: Template Matching

- Processing batches of data in parallel and the importance of GPUs
  - What are mini-batches? What are the pluses and minuses of using mini-batches?
  - **Lab 9**: Training a neural network with mini-batches

Day 2

Morning session: Deep dive into the main concepts and equations of deep learning

- Brief multivariable calculus review
  - Partial derivatives, gradients, and direction of steepest descent

- Loss and cross-entropy
  - The cross-entropy loss for classification tasks: equations, intuitions and visuals
  - Why these formulas? Where do they come from? Why are they natural?
  - **Lab 10**: Playing with the cross-entropy loss with Pytorch

- Stochastic gradient descent
  - Gradient descent and stochastic gradient descent: equations, visuals and animations
  - Why is stochastic gradient descent better than gradient descent?
  - The two most important hyper-parameters: batch size and learning rate.
  - Updating the internal parameters with the formula: “outer product between the error and the data”. Where does this formula come from?

- The Backprop algorithm
  - The visuals and the intuition behind the Backprop algorithm
  - The forward pass: multiply the data by the weight matrices
  - The backward pass: multiply the errors by the transpose of the weight matrices

Afternoon session: a) Deep dive into the implementation and training of neural net with Pytorch
b) Why does a neural network need to be deep?

- Stochastic Gradient Descent and Backprop in Pytorch
  - **Lab 11, 12, 13, 14 and 15**: Implementing from scratch a neural network for image recognition

- Understanding depth
  - Deep = Multiple layers = multiple step reasoning
  - Why do we need many layers? To answer this question, we will consider specific tasks and see how some tasks can only be solved by using multiple step reasoning.
  - Comparing the patterns detected by a one-layer network and the ones detected by a multi-layer network.

- Training a deep neural network with a GPU
  - **Lab 16**: How to use a GPU with Pytorch
  - **Lab 17**: Improving the accuracy of our neural network by adding depth

DAY 3

Morning session: a) Convolutional neural networks and application to image recognition
b) A little bit of hardware

- Understanding convolutions
  - Convolution: equations, intuition, visuals and animations
  - Max Pooling: equations, intuition, visuals and animations
  - **Lab 18**: convolutional layers and max pooling in Pytorch
  - How does backprop work in a convolutional neural network?

- Implementing a deep convolutional neural network in Pytorch
  - The VGG architecture

  - **Lab 19**: We will implement from scratch a 10-layer convolutional neural network for image recognition and obtain results which are not far from the state-of-the-art. This is a long lab!

- Developing an intuitive understanding of what happens in the various layers of a deep convolutional neural network
  - Early layers detect small patterns: edges, corners, …
  - Deep layers are much more global: they can detect a person, a car, …
  - We will look at concrete examples to get an intuitive understanding of how this hierarchy naturally occur in a deep convolutional neural network

- Hardware for deep learning
  - Efficient implementation of matrix-matrix multiplication on a GPU
  - Nvidia's Tensor Cores and Google's Tensor Processing Unit
  - Dataflow processing, low-precision arithmetic, and memory bandwidth
  - Nervana, Graphcore, Wave Computing (the next generation of AI chip?)

Afternoon session: Recurrent neural networks and applications to natural language processing

- Recurrent neural networks
  - Equations and intuition
  - Language modeling
  - Recurrent neural networks are the natural way to deal with sequential data. Why?
  - How does backprop work in a recurrent neural network?
  - Text generation

- Long Term Short Term Memory (LSTM) cells
  - Why are these equations so complicated!!!
  - Understanding gating mechanisms
  - An intuitive way to understand LSTM cells

- Implementing a recurrent neural network in Pytorch
  - **Lab 20**: We will implement from scratch a recurrent neural network with LSTM cells for a language modeling task. Then we will ask the network to complete sentences or to generate text. This is a long lab.

- Translation (e.g. Google translate)
  - The encoder and the decoder networks
  - How does backprop work in such systems?

- Speech recognition
  - Spectrogram
  - How does backprop work in such systems?

Lab Exercises:

There are numerous hands-on programming labs in this course. Among others, you will:

- Implement from scratch a 10-layer convolutional neural network for visual recognition. Train it on the cloud with a GPU, and obtain results which are not far from the state-of-the-art.
- Implement from scratch a recurrent neural network with LSTM cells for a language modeling task. Train it on the cloud with a GPU, then use the network to complete sentences or generate text.
- Observe in real time how a neural network changes its internal parameters to learn useful representations of the data.

Description of Lab Environment:

In order to have access to GPUs, we will use a modern cloud platform. At the beginning of each lab, you will simply enter an IP address in the local browser of your laptop -- you will then be connected to a Jupyter Notebook that runs on a cloud virtual machine instance with a dedicated GPU. The Jupyter Notebook will provide you with detailed instructions on how to progress through the lab.

Required Hardware:

Students must bring their laptop (Linux, Mac, or Window).

Required Prerequisites:

- Programming prerequisites:
    - Ideally you should have a good python programming background
    - But if you have a solid programing experience in a different language (e.g. C/C++/Matlab/R/Javascript), you should be fine. Just take one or two quick python tutorials before you come to the training, and make sure you understand the syntax for defining and using classes.

- Mathematical prerequisites:
    - Basic understanding of matrix-vector and matrix-matrix multiplications
    - Basic understanding of what are partial derivatives
    - We will provide quick refreshers on each of these topics, but it is preferable if you are familiar with them before to come to the training.

  Remark: the mathematical prerequisites for this training are very minimal. The reason is that deep learning algorithms, when presented correctly, are actually deceptively simple. Understanding them does not require any sophisticated mathematical theories. All you need are matrix multiplications and partial derivatives.

Course Materials:

- Downloadable PDF version of the presentation slides
- Lab exercise solutions