


Deep Learning

Chuks Okoli

Last Updated: August 7, 2024

1	Welcome to Deep Learning	2
1.1	Introduction to Deep learning	2
1.1.1	Supervised Learning with Neural Networks	3
1.1.2	Why is Deep Learning taking off?	4
1.2	Neural Network Basics	5
1.2.1	Logistic Regression as a Neural Network	5
1.3	On Cross-Referencing	9
1.4	On Math	9

WELCOME TO DEEP LEARNING



Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don't think AI (Artificial Intelligence) will transform in the next several years.

— Andrew Ng, *DeepLearning.AI*

HELLO THERE, and welcome to Deep Learning. This work is a culmination of hours of effort to create my reference for deep learning. All the explanations are in my own words but majority of the contents are based on DeepLearning.AI's specialization in [Deep Learning](#).

1.1 Introduction to Deep learning

The term, Deep Learning, refers to training Neural Networks, sometimes very large Neural Networks. In order to predict the price of a house based on its size for example, we can apply linear regression as a method for fitting a function to predict house prices. An alternative approach would be to use a simple neural network to model the relationship between house size and price.

The simple neural network consists of a single neurons, which takes an input (e.g., house size) and outputs a prediction (e.g., house price). The neuron computes a linear function of the input and applies a rectified linear unit (ReLU) activation function to ensure non-negativity. Each neuron in the network computes a function of the input features and contributes to the overall prediction. When extended to multiple features, each feature is represented by a separate neuron, and the network learns to predict the house price based on these features.

Neural networks are useful in supervised learning scenarios, where the goal is to map input features to corresponding output labels. Given enough training data, neural networks can learn complex mappings from inputs to outputs.

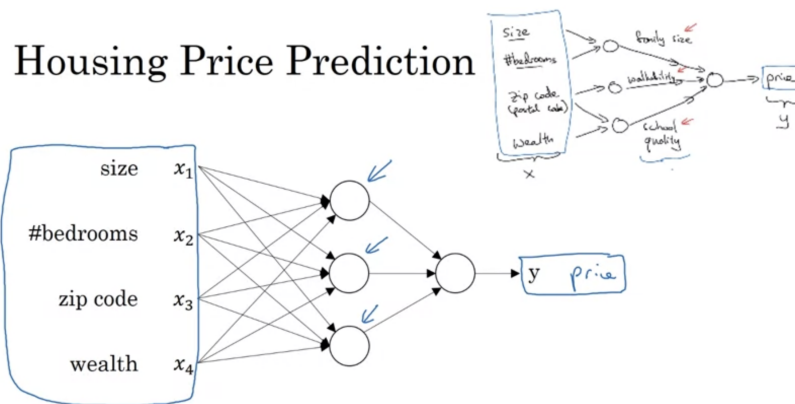


Figure 1.1: A Simple Neural Network for House Price Prediction

1.1.1 Supervised Learning with Neural Networks

Neural networks have gained a lot of attention lately for their ability to solve complex problems effectively. In supervised learning, you input data and aim to predict an output. Examples include predicting house prices or online ad clicks. Neural networks have been successful in various applications, like online advertising, computer vision, speech recognition, and machine translation. Different types of neural networks are used based on the nature of the data, such as convolutional neural networks for images and recurrent neural networks for sequential data. Structured data, like database entries, and unstructured data, like images or text, are both now interpretable by neural networks, thanks to recent advancements. While neural networks are often associated with recognizing images or text, they also excel in processing structured data, leading to improved advertising and recommendation systems. The techniques covered in this course apply to both structured and unstructured data, reflecting the versatility of neural networks in various applications.

Supervised Learning

Input(x) ↙	Output (y) ↙	Application
Home features	Price	Real Estate
Ad, user info ↙	Click on ad? (0/1)	Online Advertising
Image	Object (1,...,1000)	Photo tagging
Audio	Text transcript	Speech recognition
English	Chinese	Machine translation
Image, Radar info ↗	Position of other cars	Autonomous driving

Figure 1.2: Examples of Supervised learning

Neural Network examples

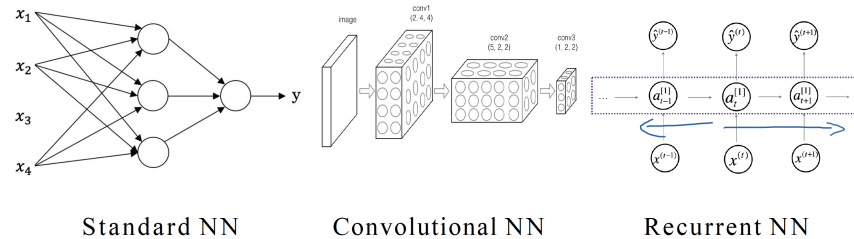


Figure 1.3: Neural network examples

1.1.2 Why is Deep Learning taking off?

The rise of deep learning has been fueled by several key factors. One major driver is the abundance of data available for training machine learning models. With the digitization of society, activities performed on digital devices generate vast amounts of data, enabling neural networks to learn from large datasets. Additionally, advancements in hardware, such as GPUs and specialized processors, have facilitated the training of large neural networks by providing faster computation speeds. Algorithmic innovations, like the adoption of the ReLU activation function, have also played a crucial role in accelerating learning processes. By reducing the time required to train models and enabling faster experimentation, these innovations have enhanced productivity and fostered rapid progress in deep learning research. Moving forward, the continued growth of digital data, advancements in hardware technology, and ongoing algorithmic research are expected to further drive improvements in deep learning capabilities. As a result, deep learning is poised to continue evolving and delivering advancements in various applications for years to come.

FUNFACT: What's drives Deep Learning

Deep Learning took off in the last few years and not before mainly because of great computing power and huge amount of data. These two are the key components for the successes of deep learning. The performance of a neural network improves with more training data.

Scale drives deep learning progress

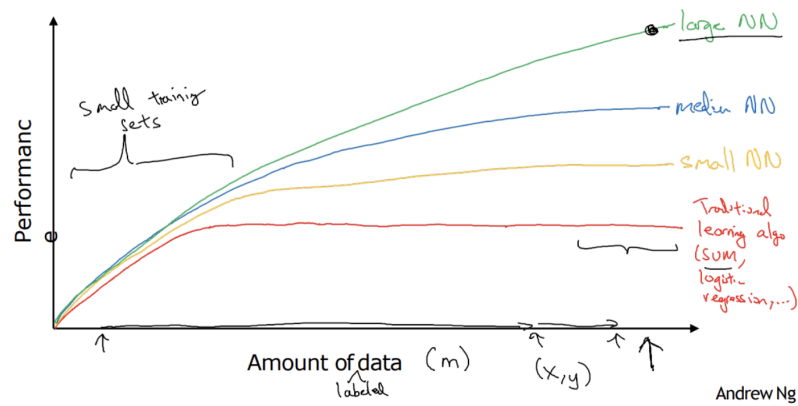


Figure 1.4: Scale drives neural networks

1.2 Neural Network Basics

1.2.1 Logistic Regression as a Neural Network

Logistic regression is an algorithm for binary classification problem. In a binary classification problem, the goal is to train a classifier for which the input is an image represented by a feature vector, x , and predicts whether the corresponding label y is 1 or 0. In this case, whether this is a cat image (1) or a non-cat image (0).



Figure 1.5: Binary classification - Cat vs Non-Cat

An image is stored in the computer in three separate matrices corresponding to the Red, Green, and Blue color channels of the image. The three matrices have the same size as the image, for example, the resolution of the cat image is 64 pixels x 64 pixels, the three matrices (RGB) are 64 x 64 each. The value in a cell represents the pixel intensity which will be used to create a feature vector of n dimension. In pattern recognition and machine learning, a feature vector represents an image. Then the classifier's job is to determine whether it contains a picture of a cat or not. To create a feature vector, x , the pixel intensity values will be "unrolled" or "reshaped" for each color. The dimension of the input feature vector x is $n = 64 * 64 * 3 = 12288$. Hence, we use $n_x = 12288$ to represent the dimensions of the feature vectors.

In binary classification, our goal is to learn a classifier that can input an image represented by this feature vector x and predict whether the corresponding label y is 1 or 0, that is, whether this is a cat image or a non-cat image.

$$x = \begin{bmatrix} 255 \\ 231 \\ 42 \\ \vdots \\ 255 \\ 134 \\ 202 \\ \vdots \\ 255 \\ 134 \\ 93 \\ \vdots \end{bmatrix} \begin{array}{l} \text{red} \\ \text{green} \\ \text{blue} \end{array}$$

Figure 1.6: Reshaped feature vector

Logistic Regression

In Logistic regression, the goal is to minimize the error between the prediction and the training data. Given an image represented by a feature vector x , the algorithm will evaluate the probability of a cat being in that image.

$$\text{Given } x, \hat{y} = P(y = 1|x), \text{ where } 0 \leq \hat{y} \leq 1 \quad (1.1)$$

The parameters used in Logistic regression are:

- The input features vector: $x \in \mathbb{R}^{n_x}$, where n_x is the number of features
- The training label: $y \in \{0, 1\}$
- The weights: $w \in \mathbb{R}^{n_x}$, where n_x is the number of features
- The threshold: $b \in \mathbb{R}$
- The output: $\hat{y} = \sigma(w^T * x + b)$
- Sigmoid function: $s = \sigma(w^T * x + b) = \sigma(z) = \frac{1}{1+e^{-z}}$

$w^T * x + b$ is a linear function ($ax + b$), but since we are looking for a probability constraint between $[0, 1]$, the sigmoid function is used. The function is bounded between $[0, 1]$ as shown in the graph above. Some observations from the graph:

1. If z is a large positive number, then $\sigma(z) = 1$
2. If z is small or large negative number, then $\sigma(z) = 0$
3. If $z = 0$, then $\sigma(z) = 0.5$

The difference between the cost function and the loss function for logistic regression is that the loss function computes the error for a single training example while the cost function is the average of the loss functions of the entire training set.

Logistic Regression cost function

$\rightarrow \hat{y}^{(i)} = \sigma(w^T x^{(i)} + b)$, where $\sigma(z^{(i)}) = \frac{1}{1+e^{-z^{(i)}}}$ $z^{(i)} = w^T x^{(i)} + b$

Given $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$, want $\hat{y}^{(i)} \approx y^{(i)}$.

Loss (error) function: $\mathcal{L}(\hat{y}, y) = \frac{1}{2} (\hat{y} - y)^2$

$\mathcal{L}(\hat{y}, y) = - (y \log \hat{y} + (1-y) \log (1-\hat{y})) \leftarrow$

If $y=1$: $\mathcal{L}(\hat{y}, y) = -\log \hat{y} \leftarrow$ want $\log \hat{y}$ large, want \hat{y} large.

If $y=0$: $\mathcal{L}(\hat{y}, y) = -\log (1-\hat{y}) \leftarrow$ want $\log (1-\hat{y})$ large ... want \hat{y} small

Cost function: $J(w, b) = \frac{1}{m} \sum_{i=1}^m \mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log \hat{y}^{(i)} + (1-y^{(i)}) \log (1-\hat{y}^{(i)})]$

Handwritten notes: $x^{(i)}$, $y^{(i)}$, $z^{(i)}$ are the i -th example. A small graph of the sigmoid function is also shown.

Figure 1.7: Logistic Regression Cost function

- Like this one,

which is wrapped in gray. I use it for notes...

- Or this one,

which is wrapped in red. I use it for fun facts or other asides...

- Or this one,

which is wrapped in blue and used for mathy stuff.

- Or this last one,

which is wrapped in green. With a title, it's used for enumerated examples (see `\extitle` and `\excounter`). Observe:

EXAMPLE 1.1: Test

This is an example. What's the answer to $2 + 2$?

ANSWER: Obviously 4, lol.

EXAMPLE 1.2: Test Again

This one will increment the counter automatically, resetting for each chapter.

- For red and blue boxes, there are custom commands for titles, too:

ONE TITLE

Like this

TWO TITLES: A Subtitle

Or this

These styles also automatically apply to theorems and claims.

Theorem 1.1 (Pythagorean Theorem). *For any right triangle with legs a, b and hypotenuse c :*

$$a^2 + b^2 = c^2 \tag{1.2}$$

Proof. This is left as an exercise to the reader. ■

Claim 1.1. *This is the greatest note template in the world.*

There are different ways to quote things, too, depending on how you want to emphasize:

This is a simple, indented quote with small letters and italics usually suitable for in-text quotations when you just want a block.

Alternatively, you can use the `\inspiration` command from the chapter heading, which leverages the `thickleftborder` frame internally, but adds a little more padding and styling (there's also just `leftborder` for a thinner variant):

■ Hello there!

1.3 On Cross-Referencing

You can reference most things—see [Theorem 1.1](#) or [\(1.2\)](#) or the [Welcome to Deep Learning](#) chapter—directly and easily as long as you give them labels. These are “built-ins.” However, you can also create a **custom term** that will be included in the index, then include references to it that link back to the original definition. Try clicking: [custom term](#). Building the index is on you, though. You can also reference by using a different term for the text: [like this](#). Sometimes it doesn’t fit the **grammatical structure** of the sentence so you can define the term one way and visualize it another way (this creates a **grammar** entry in the index). There’s also **math terms** and a way to reference them: [math terms](#) (clickable), but they do **not** show up in the index.

This is the standard way to include margin notes. There are also commands to link to source papers directly (see `\lesson`).

1.4 On Math

Most of the math stuff is just macros for specific things like the convolution operator, \otimes , probabilities, $\Pr[A|B = C]$, or big- O notation, $\mathcal{O}(n^2 \log n)$ but there’s also a convenient way to include explanations on the side of an equation:

$$\begin{array}{ll}
 1 + 1 \stackrel{?}{=} 2 & \text{first we do this} \\
 2 \stackrel{?}{=} 2 & \text{then we do this} \\
 2 = 2 & \blacksquare
 \end{array}$$

These are all in the `CustomCommands.sty` file.