Introduction to Deep Learning (CS474)

Lecture 10

Rupam

Outline

- Mechanics of Learning-PART V
    - Artificial Neural Network

Rupam

Introduction

- We've focused on a very simple regression problem that used a linear model with only one input and one output.

- Backpropagating errors to parameters and then updating those parameters by taking the gradient with respect to the loss is the same no matter what the underlying model is.

- At the core of deep learning are neural networks: mathematical entities capable of representing complicated functions through a composition of simpler functions.

- Modern artificial neural networks bear only a slight resemblance to the mechanisms of neurons in the brain.

Slide credit: E. STEVENS, L. ANTIGA, and T. VIEHMANN

Rupam

An Artificial Neuron

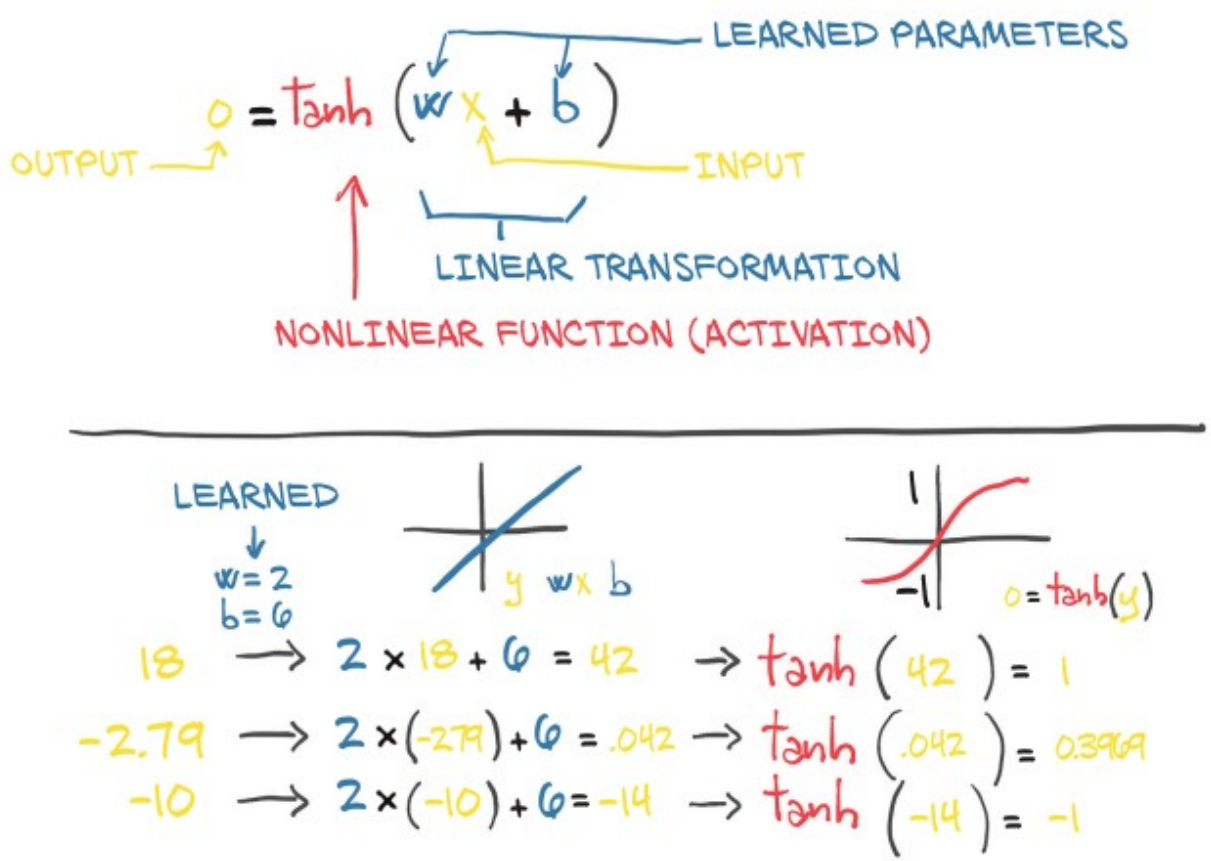

Figure 1

Image credit: E. STEVENS, L. ANTIGA, and T. VIEHMANN

Rupam

An Artificial Neuron

- It seems likely that artificial neural networks use mathematical strategies for approximating complicated functions.

- The basic building block of these complicated functions is the neuron, as illustrated in figure 1.

- At its core, it is nothing but a linear transformation of the input (for example, multiplying the input by a number [the weight] and adding a constant [the bias]) followed by the application of a fixed nonlinear function (referred to as the activation function)

Slide credit: E. STEVENS, L. ANTIGA, and T. VIEHMANN

Rupam

An Artificial Neuron

- Mathematically, we can write this out as `o = f(w * x + b)`, with **x** as our input, **w** our weight or scaling factor, and **b** as our bias or offset.

- **f** is our activation function, set to the **hyperbolic tangent**, or tanh function here.

- In general, **x** and, hence, **o** can be simple scalars, or vector-valued (meaning holding many scalar values); and similarly, **w** can be a single scalar or matrix, while **b** is a scalar or vector (the dimensionality of the inputs and weights must match, however).

Slide credit: E. STEVENS, L. ANTIGA, and T. VIEHMANN

Rupam

Multi-layer Neural Network

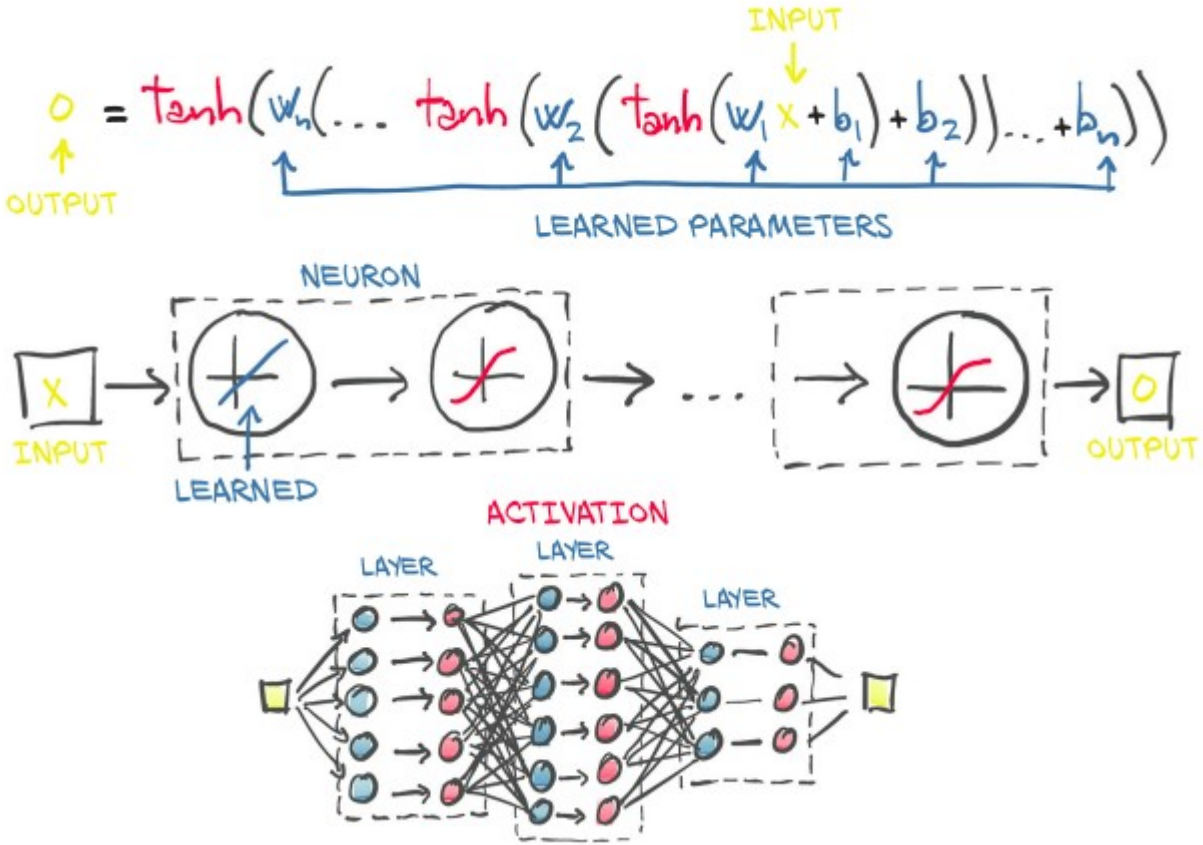

Image credit: E. STEVENS, L. ANTIGA, and T. VIEHMANN

Rupam

Understanding the Error Function

- An important difference between our earlier linear model and what we'll actually be using for deep learning is the shape of the error function.
- Our linear model and error-squared loss function had a **convex** *error curve* with a singular, clearly defined minimum.
- Neural networks **do not** have that same property of a convex error surface, even when using the same error-squared loss function!
- A big part of the reason neural networks have non-convex error surfaces is due to the activation function.

Slide credit: E. STEVENS, L. ANTIGA, and T. VIEHMANN

Rupam

Activation Function

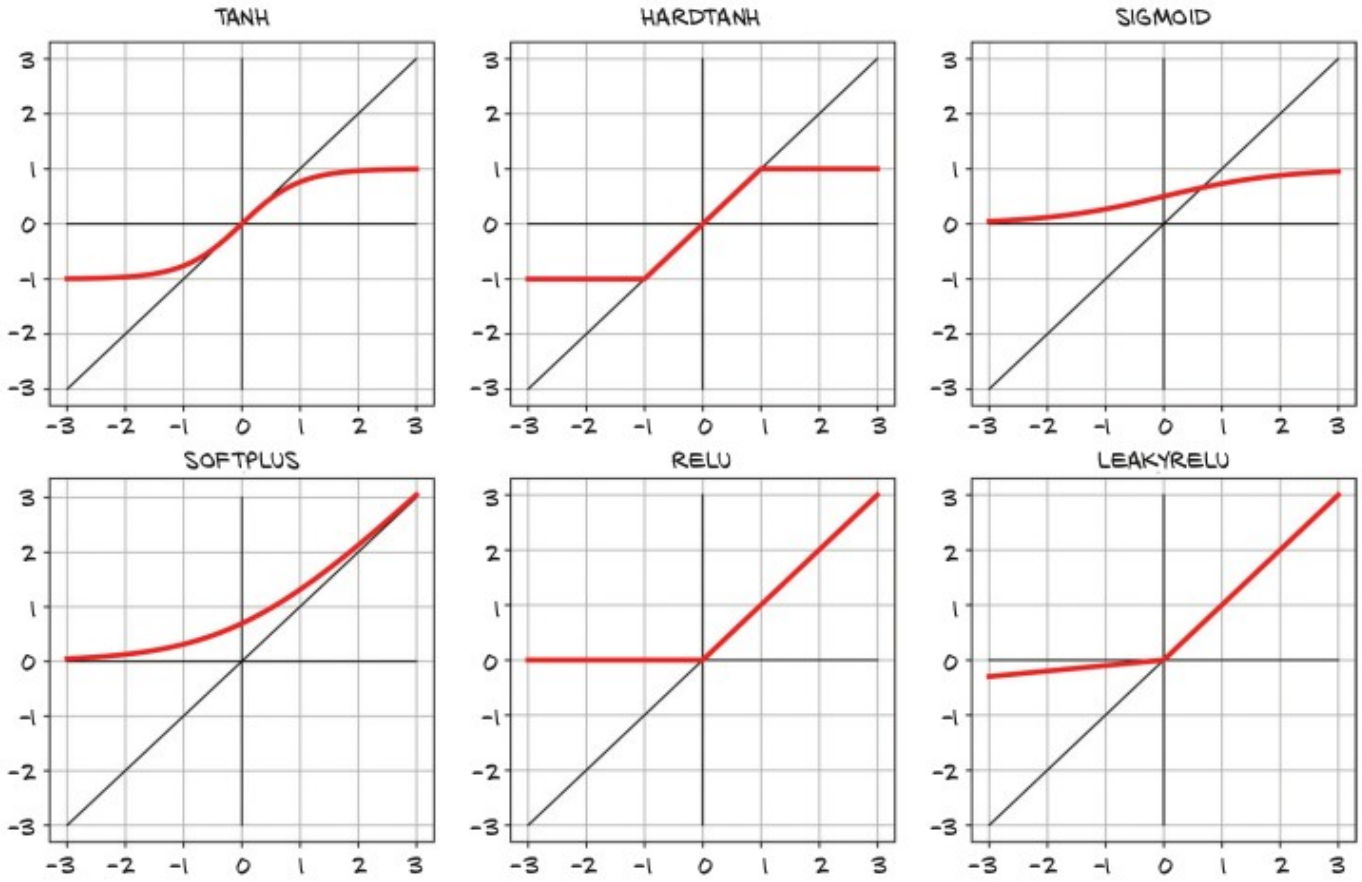

Image credit: E. STEVENS, L. ANTIGA, and T. VIEHMANN

Rupam

Activation Function

- There are quite a few activation functions, some of which are shown in the previous slide.

- In the first column, we see the smooth functions *Tanh* and *Softplus* , while the second column has "hard" versions of the activation functions to their left: *Hardtanh* and *ReLU* .

- ReLU (for rectified linear unit) deserves special note.

- The Sigmoid activation function, also known as the logistic function, was widely used in early deep learning work but has since fallen out of common use except where we explicitly want to move to the 0…1 range.

Image credit: E. STEVENS, L. ANTIGA, and T. VIEHMANN

Rupam

Activation Function

- Finally, the LeakyReLU function modifies the standard ReLU to have a small positive slope, rather than being strictly zero for negative inputs.

- The nonlinearity allows the overall network to approximate more complex functions.

- If the activation functions are differentiable, then the gradients can be computed through them.

- Without these characteristics, the network either falls back to being a linear model or becomes difficult to train.

Image credit: E. STEVENS, L. ANTIGA, and T. VIEHMANN

References

- All the contents present in the slides are taken from various online resources. Due credit is given in the respective slides.These slides are used for *academic* purposes only.