The Math Behind Neural Networks

David Brandfonbrener

Berkeley College Commonplace Society

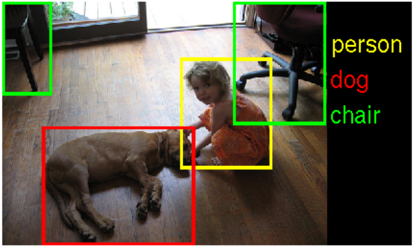
December 5, 2017

Overview

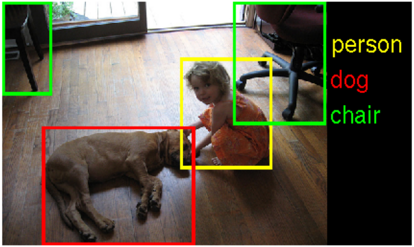
1. Background
2. Connection to research

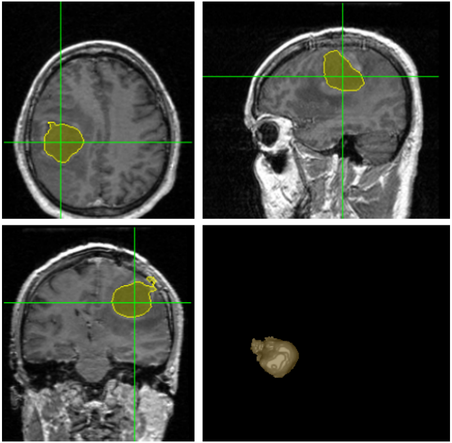
Table of Contents

Motivation

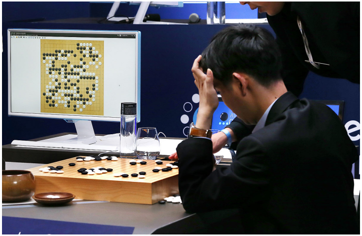


Motivation

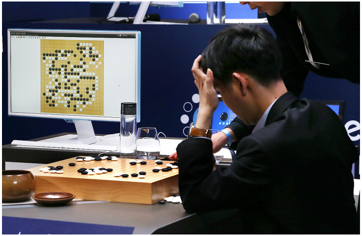




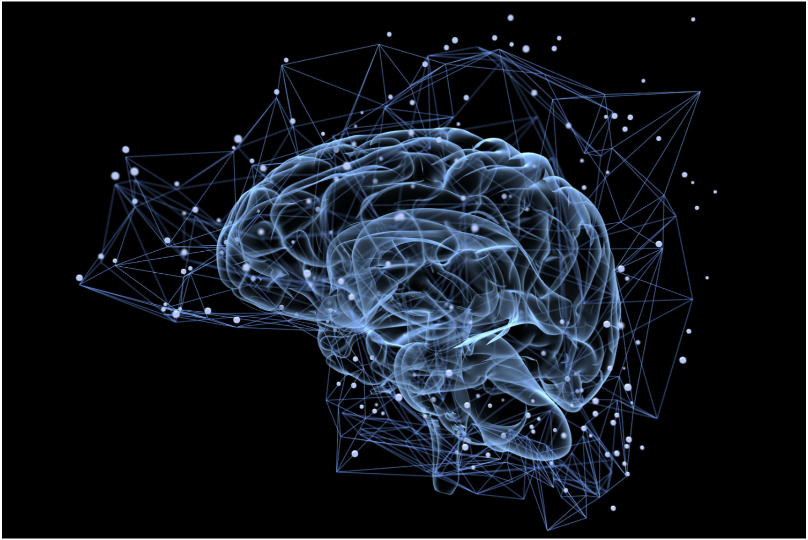
Motivation



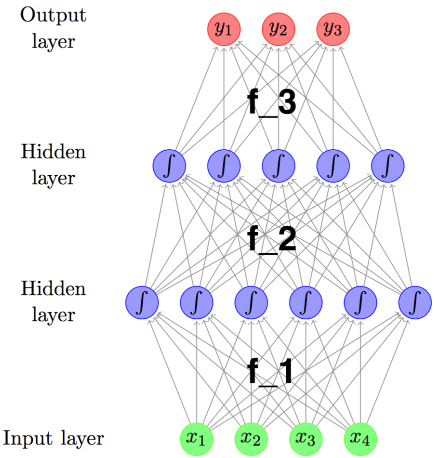
Motivation



Motivation



Model



- X inputs, Y outputs
- $f_N : X \to Y$
- $f_N(\vec{x}) = f_3(f_2(f_1(\vec{x})))$

Model (cont.)

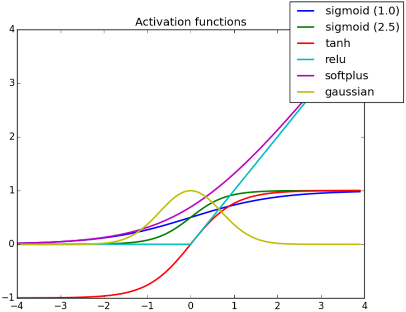
- $f_i : \mathbb{R}^d \to \mathbb{R}^k$

Model (cont.)

- $f_i : \mathbb{R}^d \to \mathbb{R}^k$
- $W_i \in \mathbb{R}^{d \times k}$

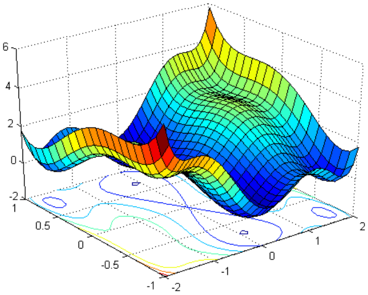
Model (cont.)

- $f_i : \mathbb{R}^d \to \mathbb{R}^k$
- $W_i \in \mathbb{R}^{d \times k}$
- σ activation function
- $f_i(\vec{x}) = \sigma(W_i \cdot \vec{x})$

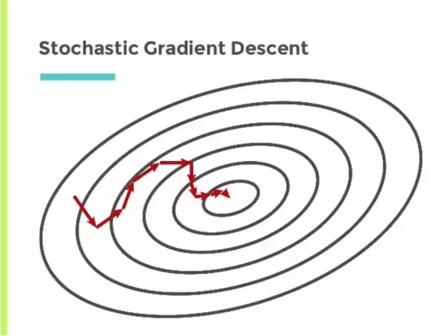


Loss Function

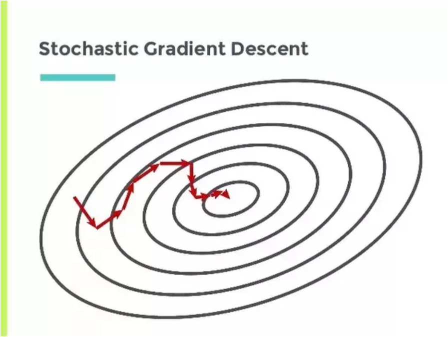
- X, Y dataset containing x_i, y_i
- W set of all W_i
- $L(W) = \sum_i (y_i - f_N(x_i))^2$



Stochastic Gradient Descent



Stochastic Gradient Descent



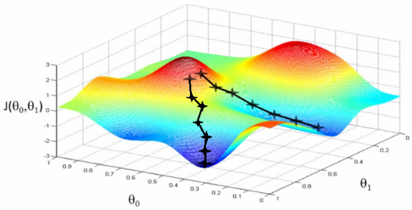
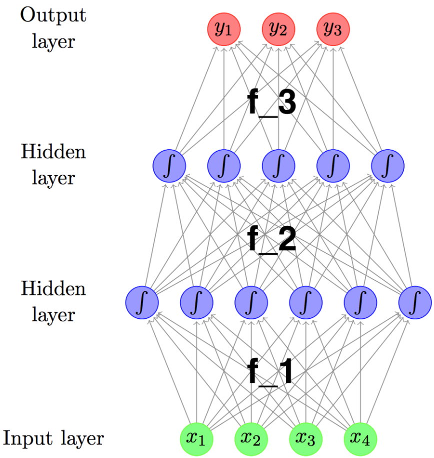


Table of Contents

Terminology



- *Depth*: number of layers
- *Width*: Number of units (neurons) per layer

Approximation Results

Universal Approximation

Any function can be approximated to arbitrary precision by a neural network with one hidden layer (depth 2).

Approximation Results

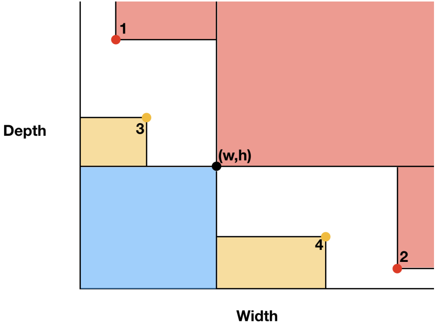
Universal Approximation

Any function can be approximated to arbitrary precision by a neural network with one hidden layer (depth 2).

Exponential Depth Efficiency

There exists a deep network (width 2, depth k) that cannot be well approximated by a shallow network with width 2^k.

Approximation Questions



- How do width and depth impact approximation capabilities?
- Is a deeper network necessarily more powerful?

Thank You!