Introduction to Neural Networks

Turing Club IISER Mohali

Motivation - Why do we need Neural Networks?

- Classification
- Data Compression
- Time series prediction
- Speech recognition
- Noise isolation
- Super resolution

and much more...

Logistic Regression

Classification - Binary and Multi-Class.

Logistic Regression: $0 \leq h_\theta(x) \leq 1$

$h_\theta(x)$ is known as the hypothesis

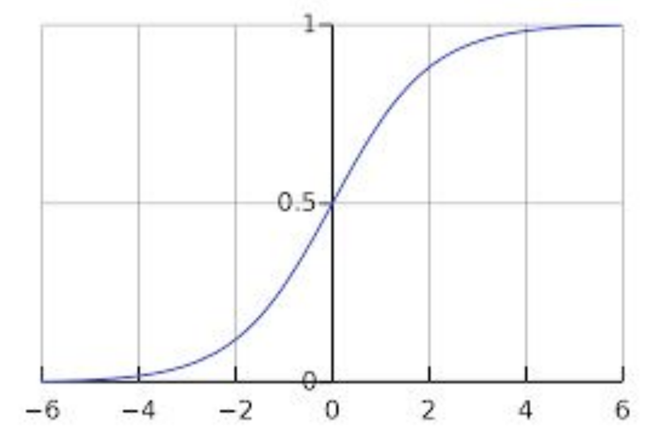
Note: Although we call it ‘regression’ it’s always used for classification

Logistic Regression

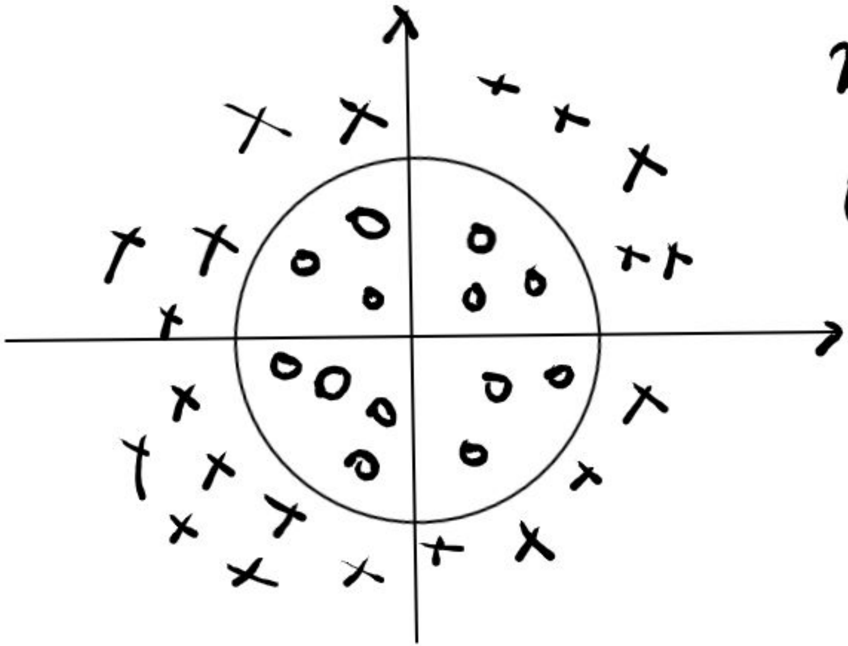
$$h_\theta(x) = g(\theta^T x)$$

$$g(z) = \frac{1}{1 + e^{-z}} \quad \text{(sigmoid)}$$

$$\Rightarrow h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$



Decision Boundary



$$x = (1 \ x_1 \ x_2 \ x_1^2 \ x_2^2)^T$$

$$\theta^T = (-1 \ 0 \ 0 \ 1 \ 1)$$

$$y = 1 \text{ if } h_\theta(x) \geq 0.5$$

$$y = 0 \text{ if } h_\theta(x) < 0.5$$

Cost Function

$$\mathrm{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) ; & y = 1 \\ -\log(1 - h_\theta(x)), & y = 0 \end{cases}$$

$$\mathrm{Cost} = -\frac{1}{m}\left[\sum_{i=1}^{m} y^i \log h_i + (1 - y^i) \log(1 - h_i)\right]$$

$$\min_\theta \mathrm{Cost}(\theta)$$

Gradient Descent

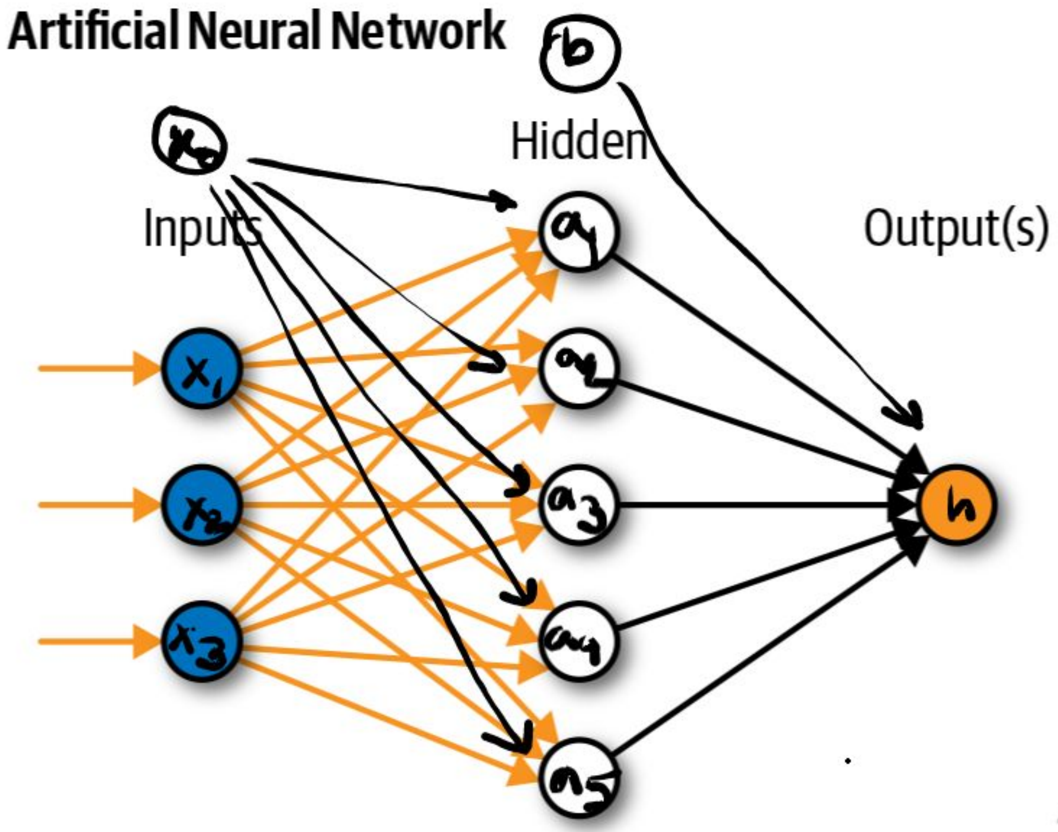
$$\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j}$$

$$J(\theta) = \text{Cost}$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i}^{m} \left(h_\theta(x^i) - y^i\right) x_j^i$$

Neural Network

Logistic Network (Binary Classification)



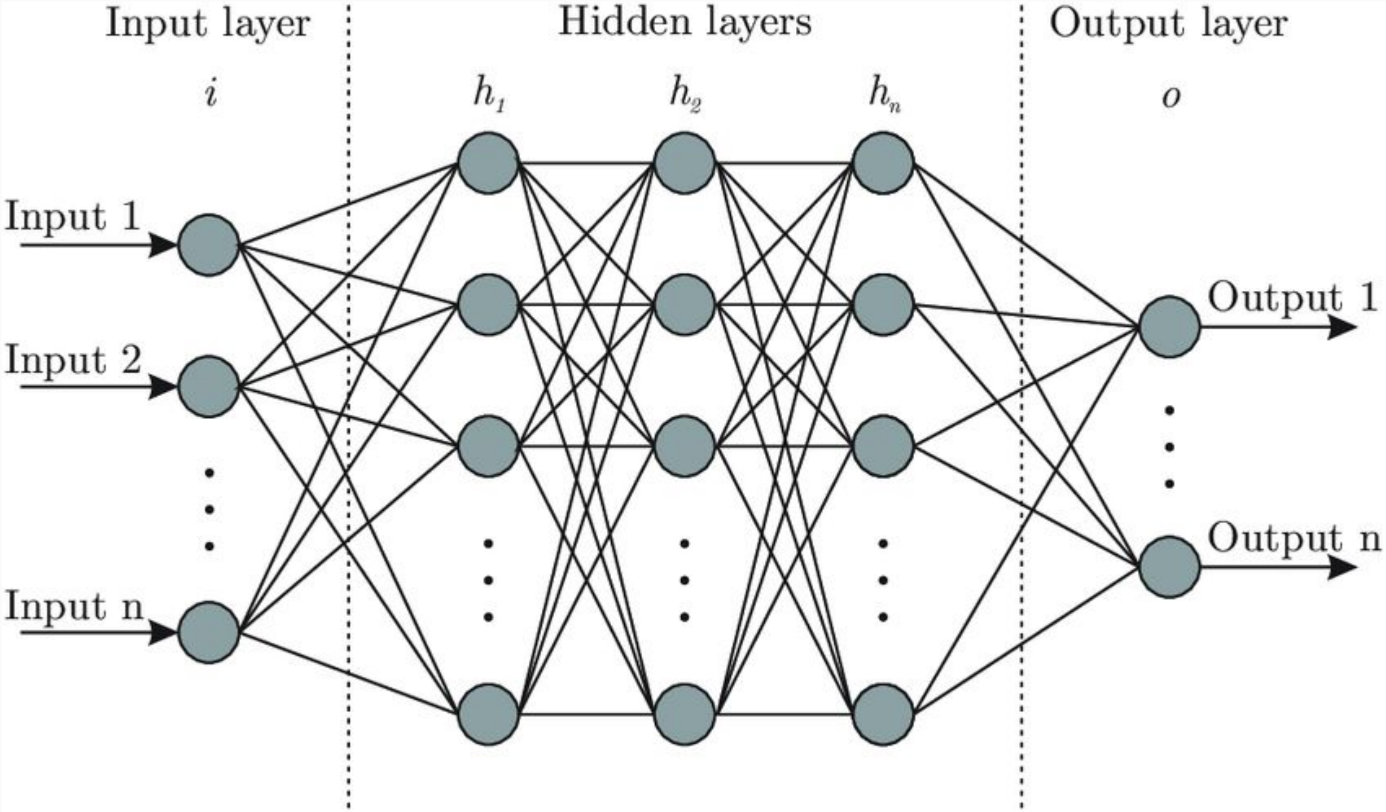
Notation

a^j_i = activation of unit i in layer j

Θ^j = matrix of weights controlling function mapping from layer j to layer j+1

$z^n = (\Theta^{n-1})^T.a^{n-1}$
$a^n = g(z^n)$

Multi-class Network



NEAT Algorithm

Neuroevolution of Augmenting Topologies

By Kenneth O. Stanley and Risto Miikkulainen
Department of Computer Sciences
The University of Texas at Austin
http://nn.cs.utexas.edu/downloads/papers/stanley.ec02.pdf

Challenges and Solutions

“(1) Is there a genetic representation that allows disparate topologies to crossover in a meaningful way? Our solution is to use historical markings to line up genes with the same origin.

(2)How can topological innovation that needs a few generations to optimize be protected so that it does not disappear from the population prematurely? Our solution is to separate each innovation into a different species.

(3) How can topologies be minimized throughout evolution without the need for a specially contrived fitness function that measures complexity? Our solution is to start from a minimal structure and grow only when necessary.”

Genetic Encoding

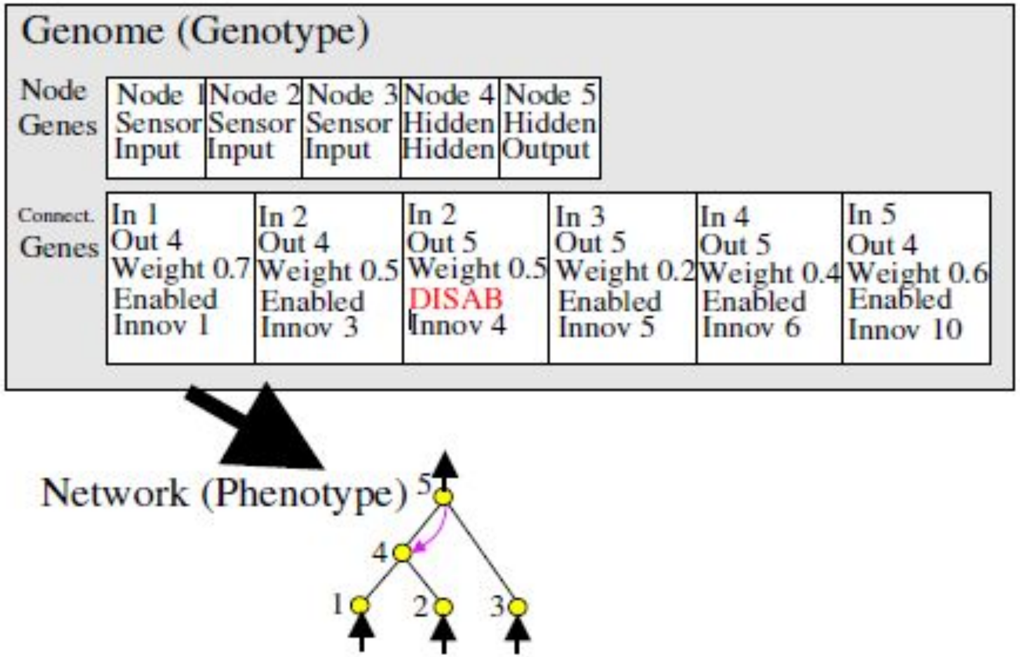


Fig. 1. **A genotype to phenotype mapping example.** The third gene is disabled, so the connection that it specifies (between nodes 2 and 5) is not expressed in the phenotype.

Genetic Encoding

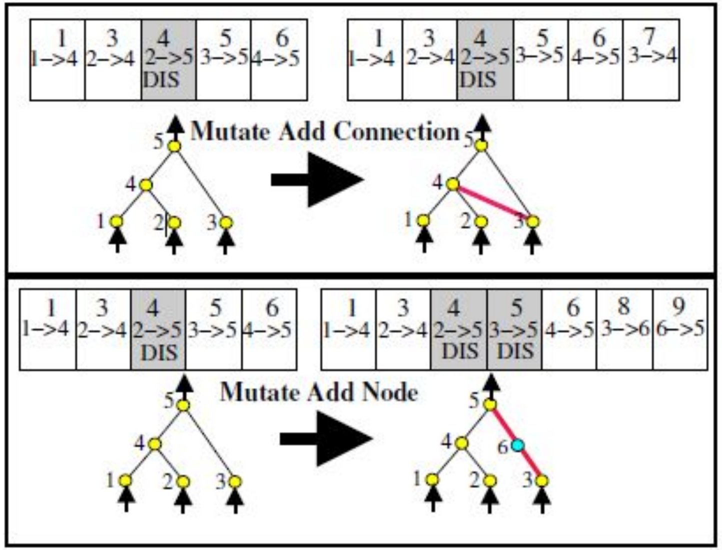


Fig. 2. **The two types of structural mutation in NEAT.** Both types, adding a connection and adding a node, are illustrated with the genes above their phenotypes. The top number in each genome is the *innovation number* of that gene. These numbers identify the original historical ancestor of each gene, making it possible to find matching genes during crossover. New genes are assigned new increasingly higher numbers.

Crossover

“In order to perform crossover, the system must be able to tell which genes match up between any individuals in the population. The key observation is that two genes that have the same historical origin represent the same structure (although possibly with different weights), since they were both derived from the same ancestral gene from some point in the past. Thus, all a system needs to do to know which genes line up with which is to keep track of the historical origin of every gene in the system.”

Crossover

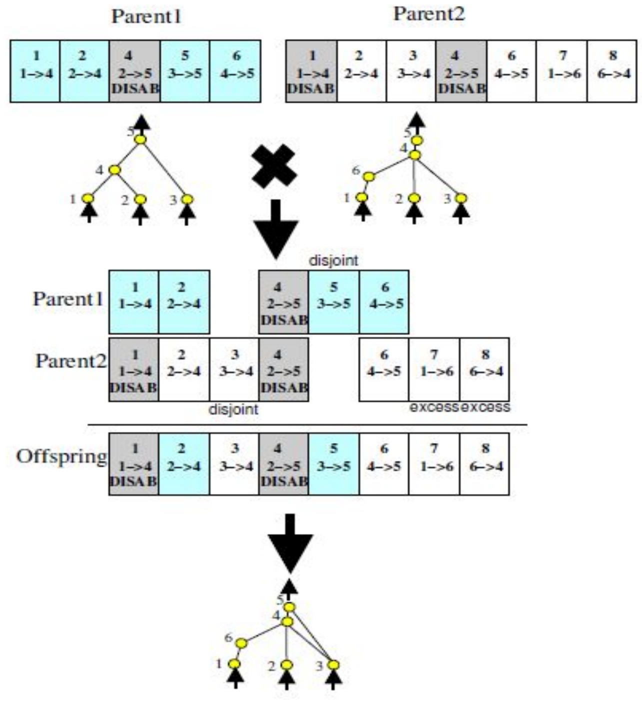


Fig. 3. **Matching up genomes for different network topologies using innovation numbers.** Although Parent 1 and Parent 2 look different, their innovation numbers (shown at the top of each gene) tell us which genes match up with which without the need for topological analysis.

Speciation

"Adding new structure to a network usually initially reduces tness. However, NEAT speciates the population, so that individuals compete primarily within their own niches instead of with the population at large."

Speciation

Compatibility distance

$$\delta = \frac{c_1 E}{N} + \frac{c_2 D}{N} + c_3 \cdot \overline{W}.$$

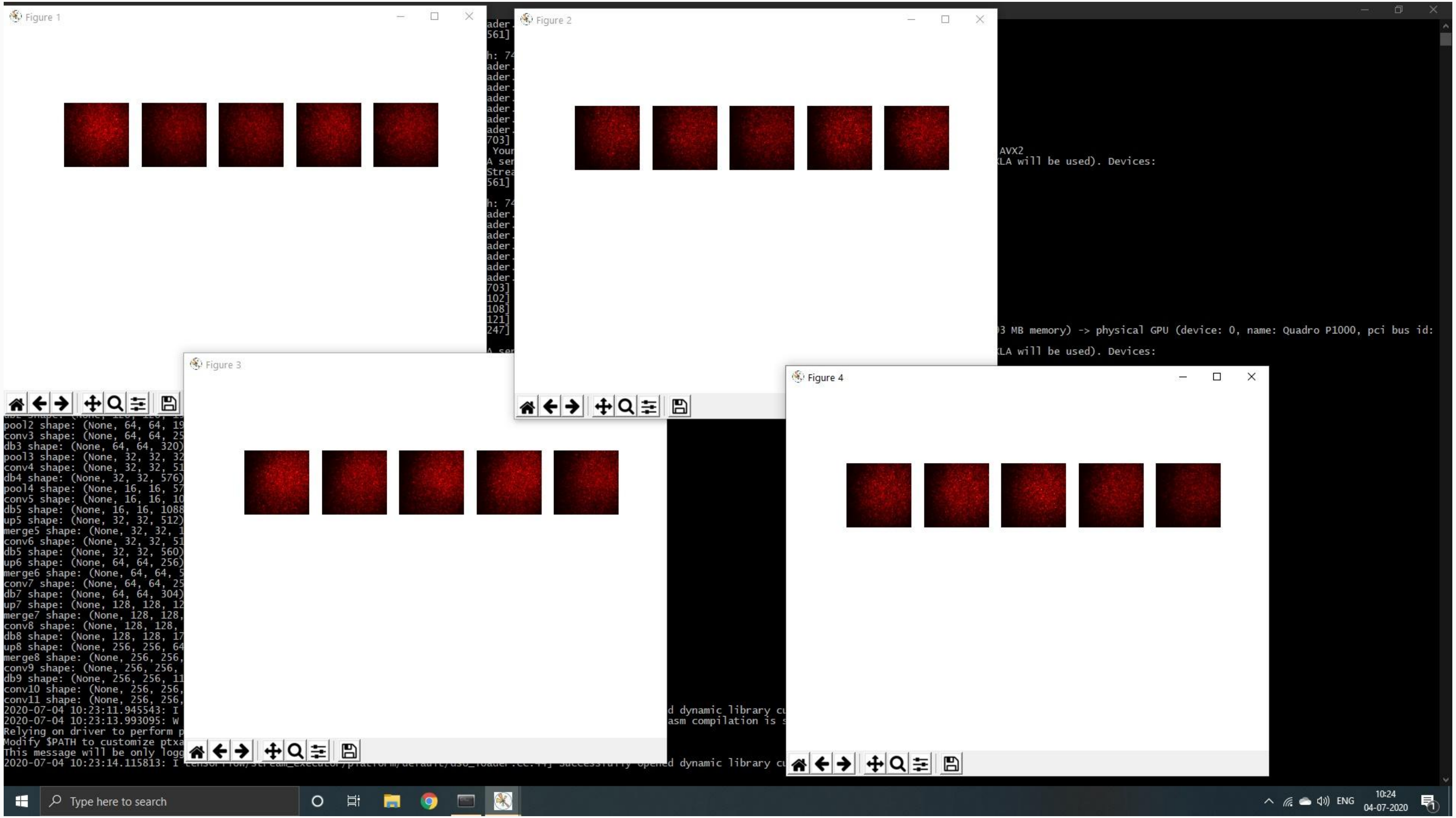
Population

$$N'_j = \frac{\sum_{i=1}^{N_j} f_{ij}}{\overline{f}},$$


Flappy Bird

Beyond Humans

Deep Speckle Correlation



Deep Speckle Correlation