

Neural Networks

Sumohana S. Channappayya

Review

- Machine Learning
 - **Supervised**
 - Unsupervised
 - Reinforcement

Review: Supervised Learning

Input (\mathbf{x})	Label (y)
	Cat
	Bird
	Dog
	Fish
	?

Review: Supervised Learning

- Given a set of **input points** and corresponding **ground truth labels**: $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$
 - Discrete labels: **classification**
 - Continuous labels: **regression**
- Learn a **parameterised (θ) functional relationship** to **estimate label** from input: $\hat{y} = f(\mathbf{x}; \theta)$
- Makes use of **training data** to learn parameters $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$

Review: Supervised Learning

- How to learn parameters θ ?
 - Quantify error: $d(y, \hat{y})$
 - Find **error** between estimate \hat{y} and ground truth over **training data** points: $R(\theta) = \sum_{i=1}^n d(y_i, \hat{y}_i)$
 - Find parameters that **minimise error**
- How **good** is our **model**?
 - Evaluate model's performance on test data $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$

Review: Supervised Learning

- Models covered so far:
 - Linear regression
 - k-NN
 - Naive Bayes
 - Logistic regression
 - Support Vector Machines (SVM)

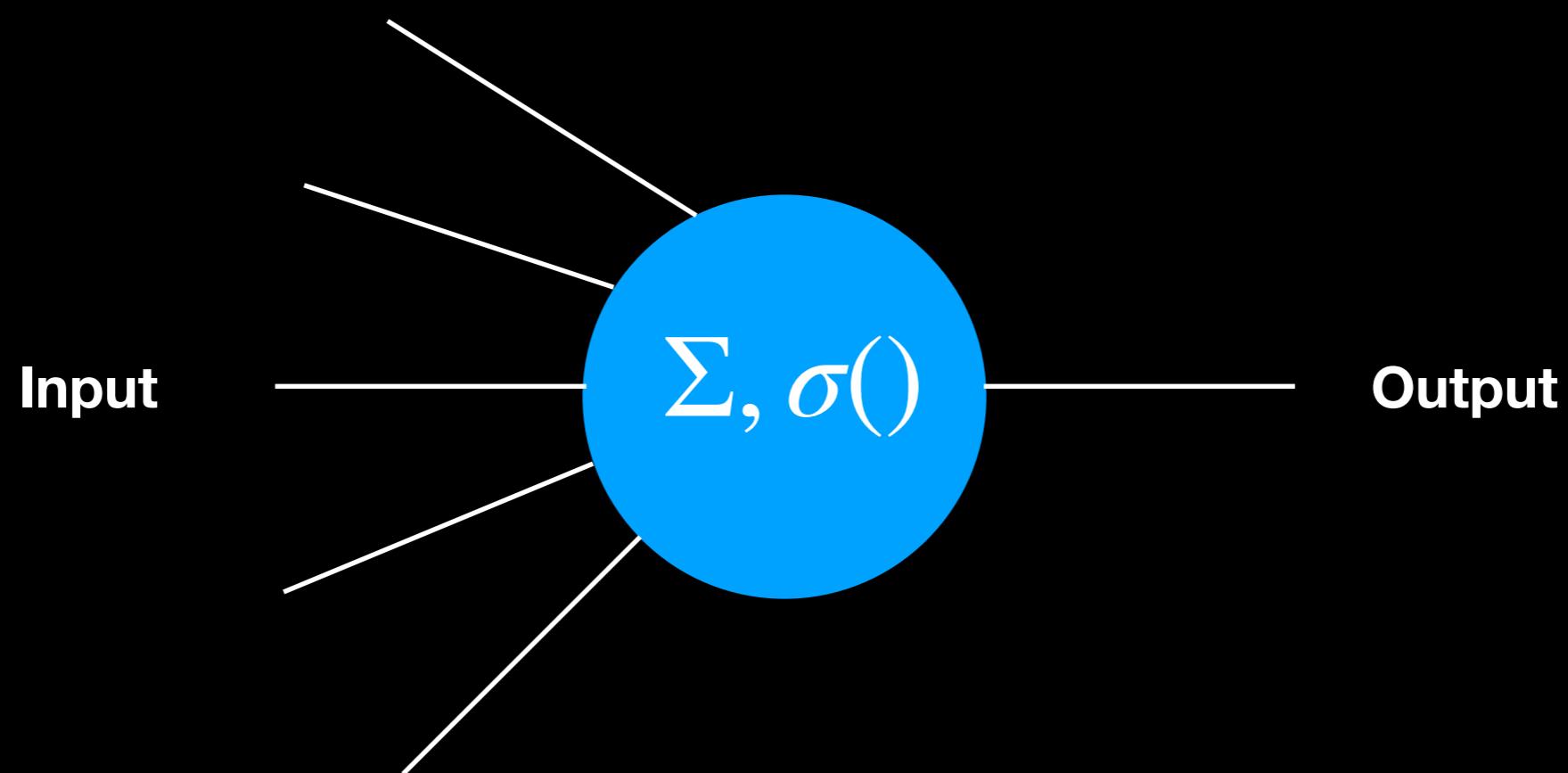
Motivation for Neural Networks

- Modelling the neuron
- Modelling **nonlinear** relation between $\{\mathbf{x}_i\}$ and $\{y_i\}$



Input (\mathbf{x})		Label (y)
0	0	0
0	1	1
1	0	1
1	1	0

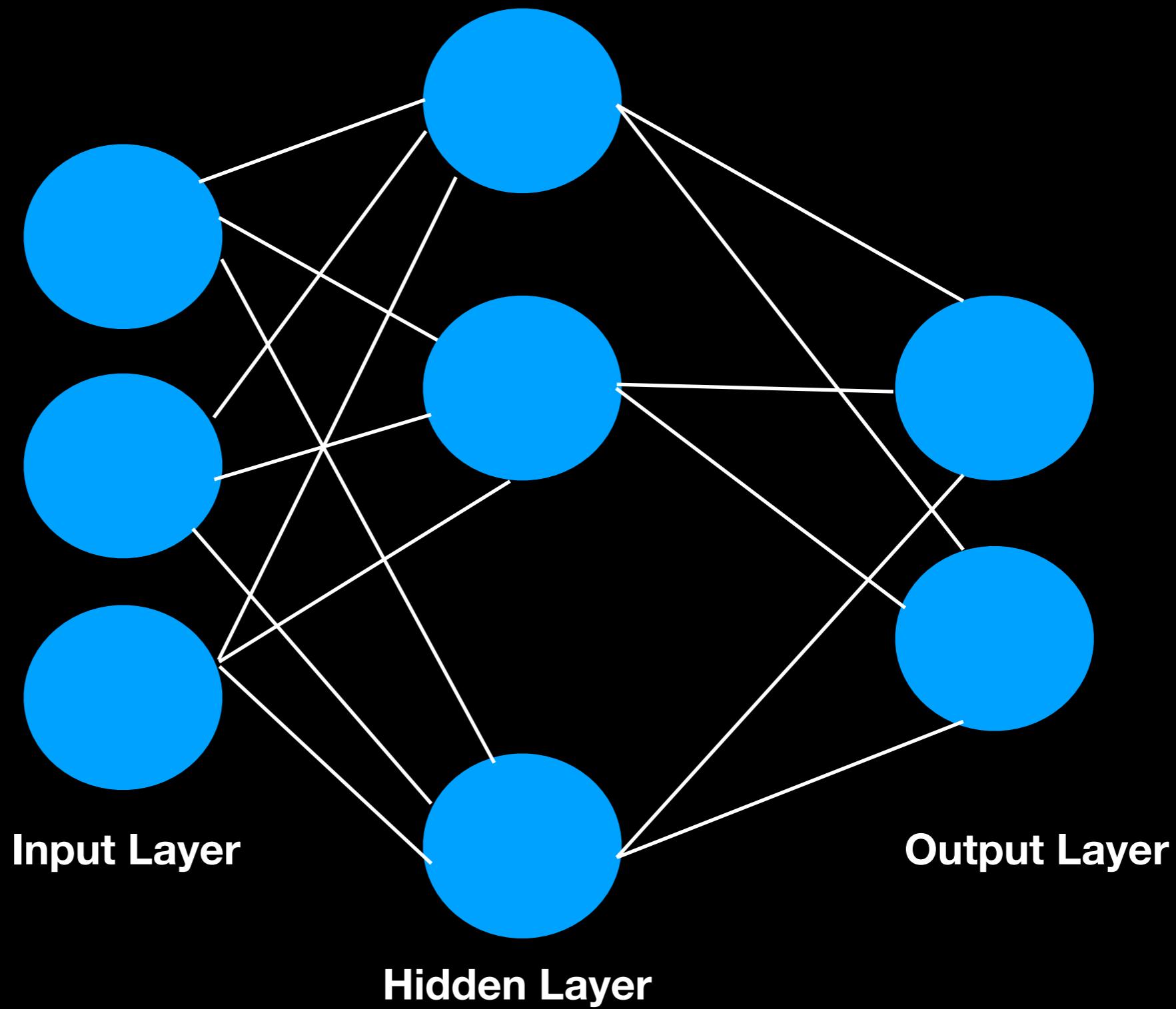
A Neuron



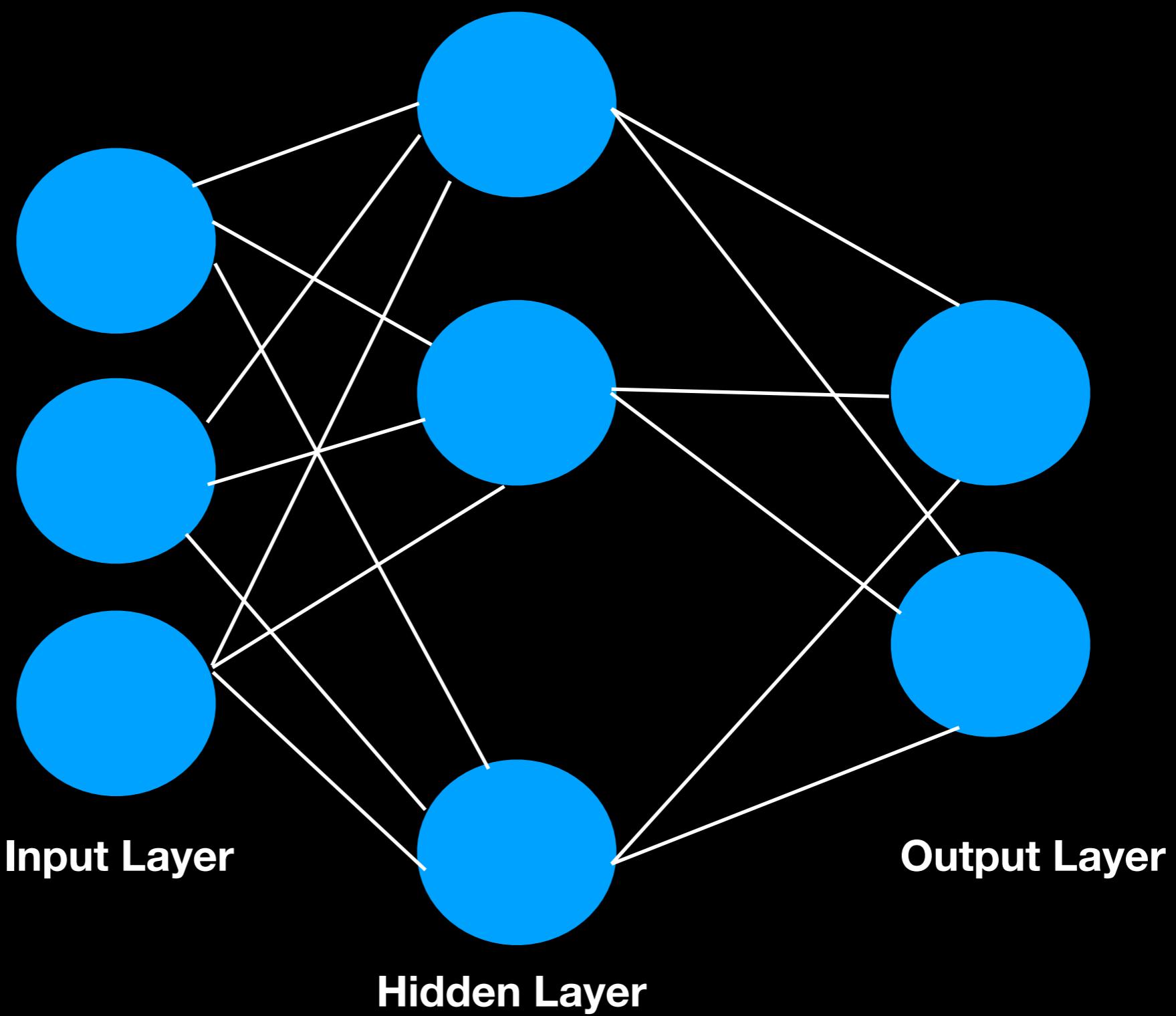
Key function: $\sigma()$

- Recall motivation: **nonlinear** input output relationship
- Properties of $\sigma()$
 - Nonlinear
 - Continuous
 - Differentiable?
 - Examples: tanh, sigmoid, ReLU etc.

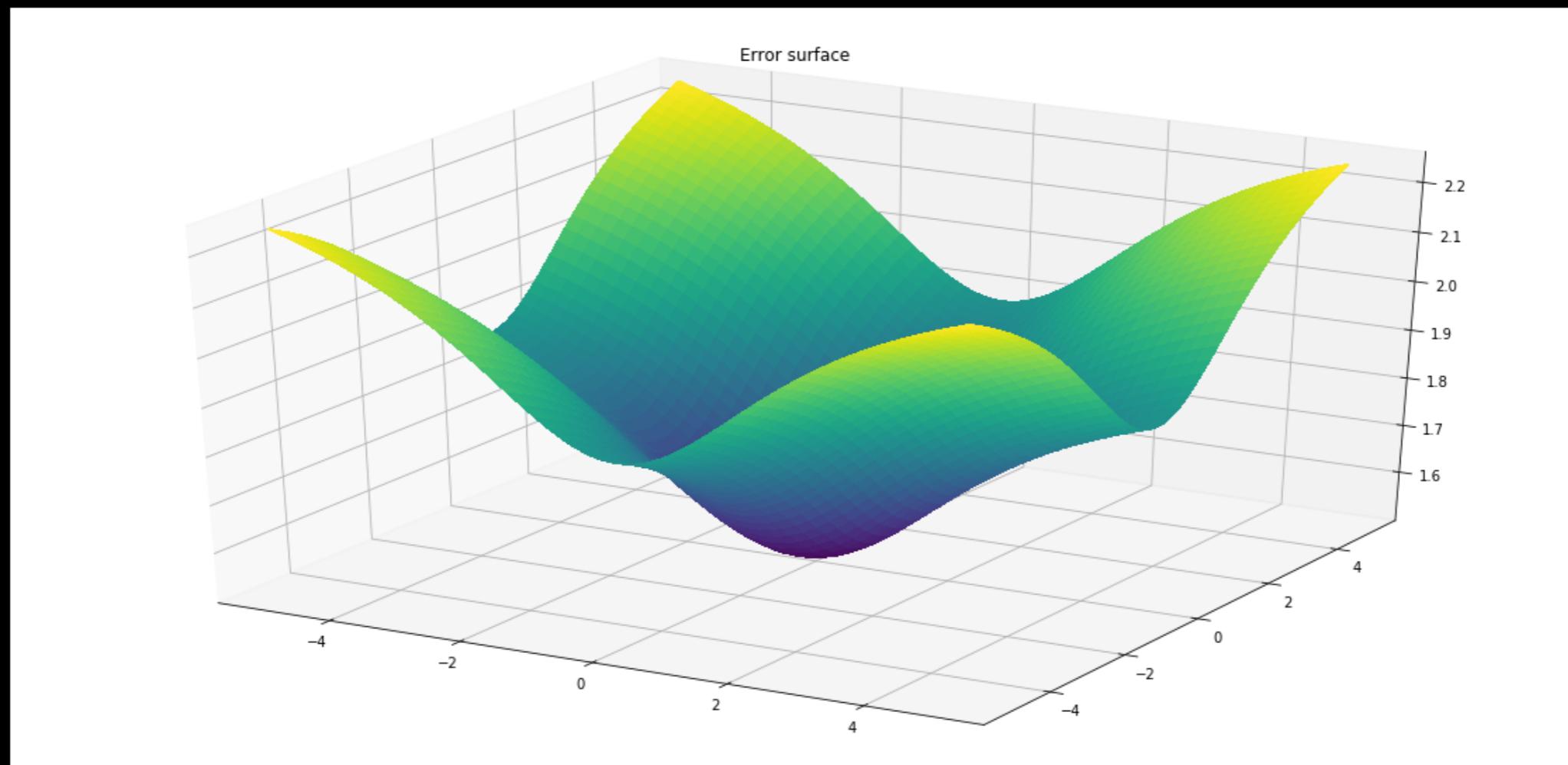
A Neural Network



Feedforward Neural Network



Error Surface



Minimising Error: $\frac{\partial R(\theta)}{\partial \theta_i}$

- Find local minimum
- Iterative procedure
- Gradient descent
- Let's derive the gradient update expression

Minimising Error: $\frac{\partial R(\theta)}{\partial \theta_i}$

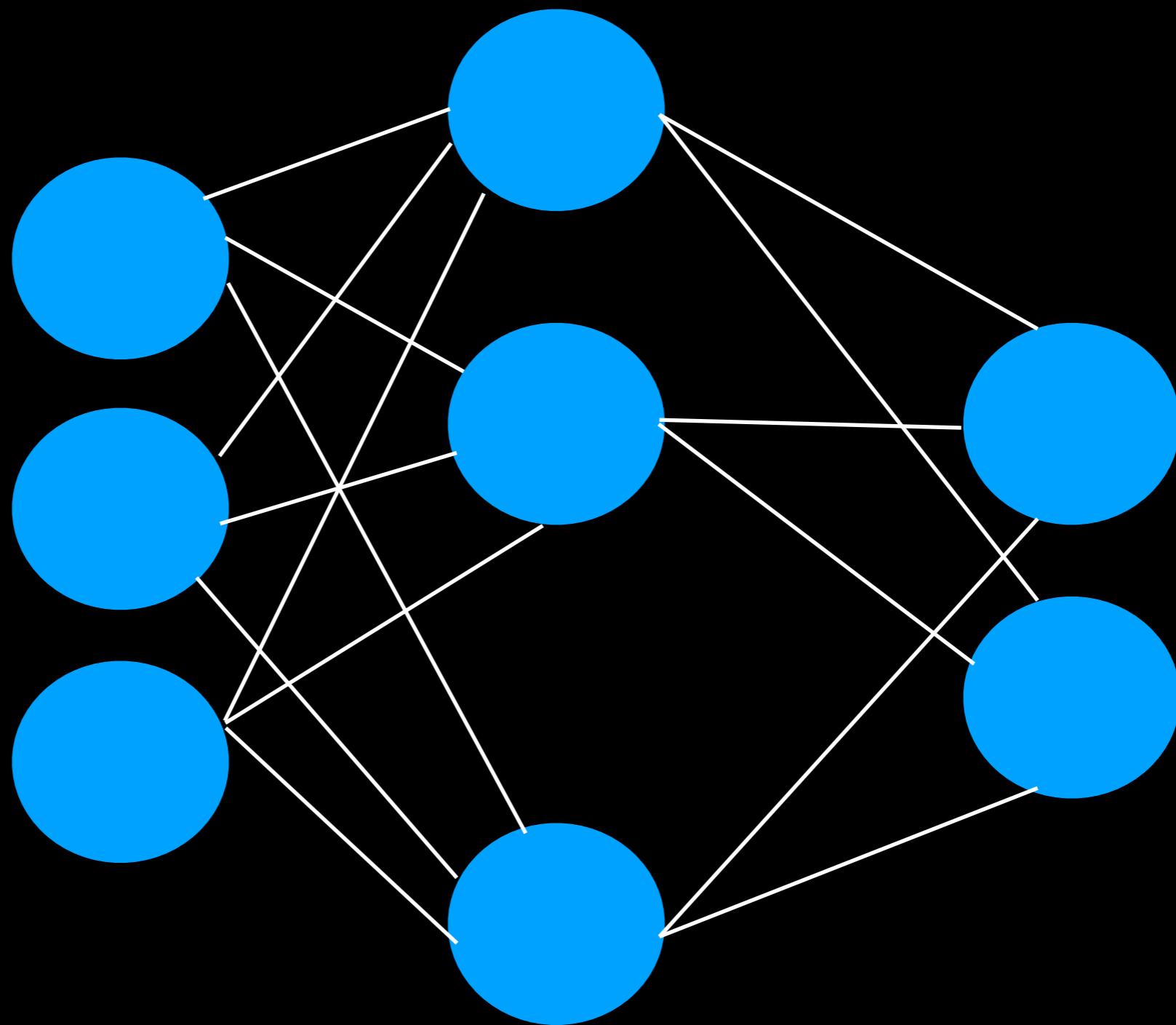
- Derivation:

Minimising Error: $\frac{\partial R(\theta)}{\partial \theta_i}$

- Derivation:

Back propagation

Back propagation



Challenges

- Initialization
- Overfitting
- Input range
- Network architecture
- Multiple minima