Boston Housing Data Example

crim

 per capita crime rate by town

zn

 proportion of residential land zoned for lots over 25,000 sq.ft.

indus

 proportion of non-retail business acres per town

chas

 Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)

nox

 nitrogen oxides concentration (parts per 10 million)

rm

 average number of rooms per dwelling

age

 proportion of owner-occupied units built prior to 1940

dis

 weighted mean of distances to five Boston employment centres

rad

 index of accessibility to radial highways

tax

 full-value property-tax rate per $10,000

ptratio

 pupil-teacher ratio by town

medv

 median value of owner-occupied homes in $1000

**Neural Networks**

Artificial neural networks are forecasting methods that are based on simple mathematical models of the brain. They allow complex nonlinear relationships between the response variable and its predictors.

**Neural network architecture**

A neural network can be thought of as a network of “neurons” organized in layers. The predictors (or inputs) form the bottom layer, and the forecasts (or outputs) form the top layer. There may be intermediate layers containing “hidden neurons”.

The simplest networks contain no hidden layers and are equivalent to linear regression.

Example: the neural network version of a linear regression with four predictors. The coefficients attached to these predictors are called “weights”. The forecasts are obtained by a linear combination of the inputs. The weights are selected in the neural network framework using a “learning algorithm” that minimizes a “cost function” such as MSE – Mean Squared Error = SSE/number\_data\_points.

In this simple example, linear regression will be much more efficient method for training the model.



Once we add an intermediate layer with hidden neurons, the neural network becomes non-linear.

This is known as a *multilayer feed-forward network* where each layer of nodes receives inputs from the previous layers. The outputs of nodes in one layer are inputs to the next layer. The inputs to each node are combined using a weighted linear combination. The result is then modified by a nonlinear function before being output.

