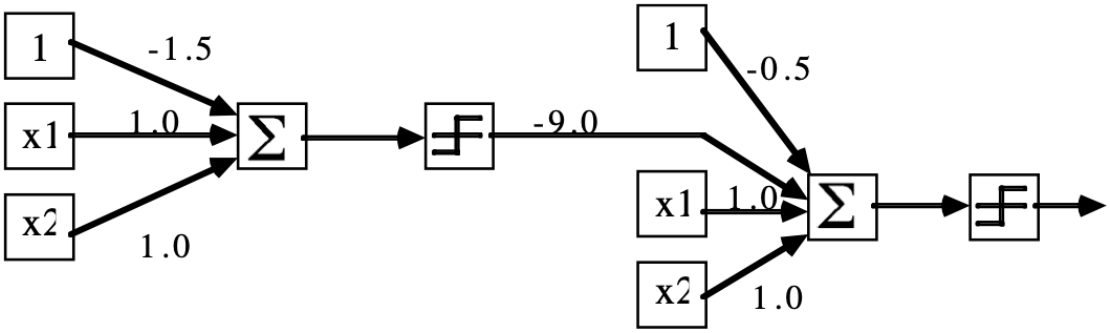
Neural Networks Tutorial

1. (a) Even though the XOR Problem is linearly unseparable, the following arrangement of two Perceptrons is able to classify the XOR inputs correctly. Show that this is true. Assume that the transfer function for each neuron is a step function that outputs 1 for positive and 0 for negative input.



(b) The -9.0 weight linking the output of the first neuron to the input of the second is much larger than is necessary. What is the minimum value that this can have and still produce the correct results?

2. The Neural Network implementation in `sklearn` provides the `alpha` parameter to control overfitting.
Overfitting can also be managed by controlling the model complexity, by reducing the number of layers and by reducing the number of units in each layer.
   - Use the Diabetes data for this analysis.
   - Produce a graph of training and test set accuracy for different numbers of units in a neural network with a single hidden layer.
   - Run the evaluation from 2 to 40 units in steps of 2.
   - Fix `alpha` at 0.15.
   - Use the graphing code from the `14 Neural Networks` notebook to plot your results.

3. Use the grid search facility in `sklearn` to find good values for `alpha` and the hidden layer size for the Neural Net.
   - A grid-search example is available in notebook 13 Grid Search.
   - You may need to run the grid-search a few times with different grids to home in on a good solution.

4. The data in the table below is from a data mining application from insurance sales that attempts to identify customers likely to buy caravan insurance. In its present format the data has 86 fields, the 86th field being the feature we wish to be able to predict (assuming we want to facilitate this data-mining exercise!). Not all the data fields are shown here – merely a representative sample. Each data record describes a household; the first feature is an enumerated type, the second is a number between 1 and 10, etc.

   The training data is tagged with whether a household has a caravan insurance policy or not. The objective is to be able to take an untagged record in the same format and predict if it is a good prospect for selling caravan insurance. In this way people who have no prospect of buying caravan insurance will be spared the junk mail.

1 **Customer Subtype**
- 1 High Income, expensive child
- 2 Very Important Provincials
- 3 High status seniors
- 4 Affluent senior apartments
- 5 Mixed seniors
- 6 Career and childcare
- 7 Dinki's (double income no kids)
- … <41 codes in all>

2 **Number of houses** 1 – 10
3 **Avg size household** 1 – 6
4 **Avg age**
- 1 20-30 years
- 2 30-40 years
- 3 40-50 years
- 4 50-60 years
- 5 60-70 years
- 6 70-80 years

5 **Customer main type**
- 1 Successful hedonists
- 2 Driven Growers
- 3 Average Family
- 4 Career Loners
- 5 Living well
- 6 Cruising Seniors
- 7 Retired and Religeous
- 8 Family with grown ups
- 9 Conservative families
- 10 Farmers

6 **Married**
7 **Living together**
8 **Singles**
9 **Household without children**
10 **Household with children**
11 **High level education**
12 **Medium level education**
13 **Lower level education**
14 **Social class A**
15 **Social class B1**
16 **Social class B2**
17 **Social class C**
18 **Social class D**
19 **Rented house**
20 **Home owners**
21 **1 car**
22 **2 cars**
23 **No car**
24 **Average income**
25 **Contribution car policies**
26 **Contribution motorcycle/scooter policies**
27 **Contribution trailer policies**
28 **Contribution tractor policies**
29 **Contribution life insurances**
… <several other similar features>

86 **Number of caravan policies 0 – 1**

(a) Describe in outline the architecture of a feedforward neural network that might be trained to predict whether a given customer is a candidate for caravan insurance.
(b) Show how the features presented might be mapped onto the inputs of the neural network.
(c) Examine the data and make proposals for compressing the number of inputs to the network.