Machine Learning

V. Adamchik | CSCI 567 | Fall 2019

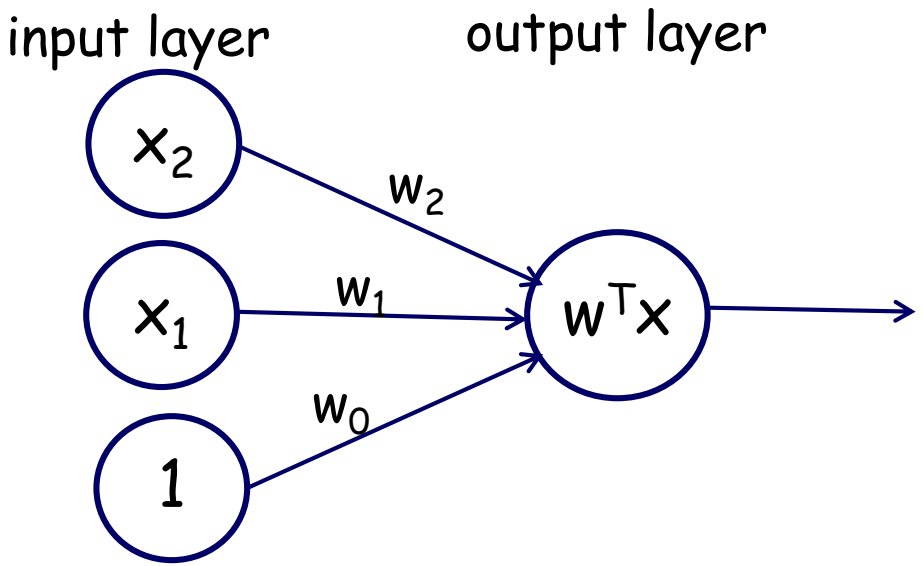
Discussion Set 6.1 | University of Southern California

Neural Nets

Problem 1

A neural network with one hidden layer and with a linear activation function can perfectly fit an XOR function.

What is the minimum number of hidden neurons required to fit the XOR function?



Problem 2

A neural network with one hidden layer and with a non-linear activation function can perfectly fit an XOR function.

What is the minimum number of hidden neurons required to fit the XOR function?

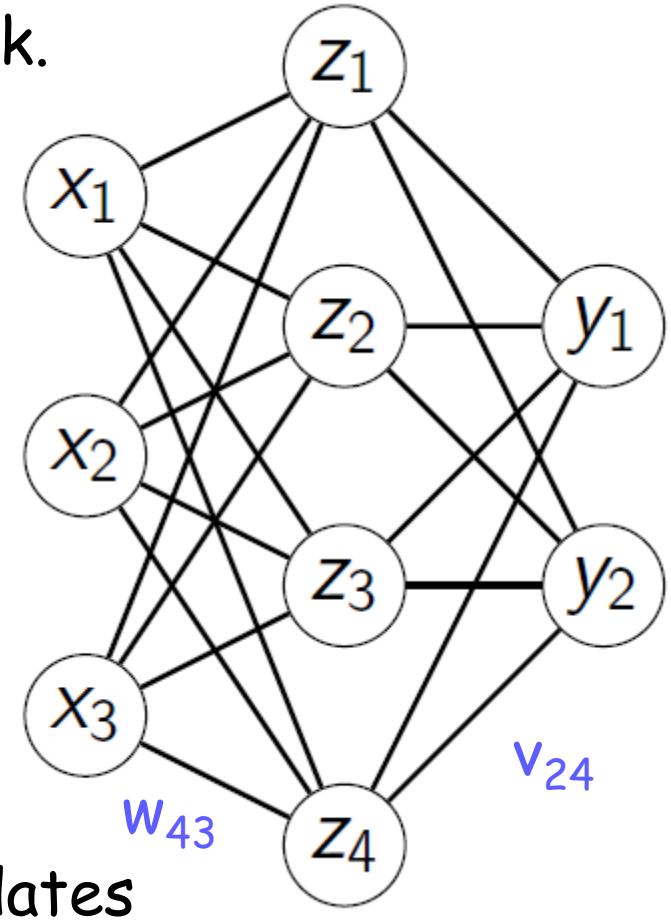
Problem 3

Consider the following neural network.

$$z_k = tanh(\sum_{i=1}^{3} w_{k\,i}\, x_i)$$

$$y_j = \sum_{k=1}^{4} v_{jk}\ z_k$$

$$L(y, y*) = \frac{1}{2}\left((y_1 - y_1^*)^2 + (y_2 - y_2^*)^2\right)$$



Write down the backpropagation updates (gradient) for estimation of w_{ki} and v_{jk}.

$$\frac{\partial}{\partial x}\tanh(x) = 1 - \tanh^2(x)$$

Problem 5

Suppose a convolution layer takes a 4×6×3 image as input and outputs a 3×4×6 tensor. Which of the following is a possible configuration of this layer?

(A) One 2×3×6 filter, stride 1, no zero-padding.
(B) Six 2×3×3 filters, stride 1, no zero-padding.
(C) Six 3×4×3 filters, stride 2, no zero-padding.
(D) Six 3×4×3 filters, stride 1, 1 zero-padding.

Problem 6

Consider the following CNN. An 8×8×3 image input, followed by a convolution layer with 2 filters of size 2×2 (stride 1, no zero-padding), then another convolution layer with 4 filters of size 3×3 (stride 2, no zero-padding), and finally a max-pooling layer with a 2×2 filter (stride 1, no zero-padding).

(1) How many parameters do we need to learn for this network?

(2) What is the picture final dimension?