Homework 2

Introduction to Machine Learning
Fall 2018
Instructor: Anna Choromanska

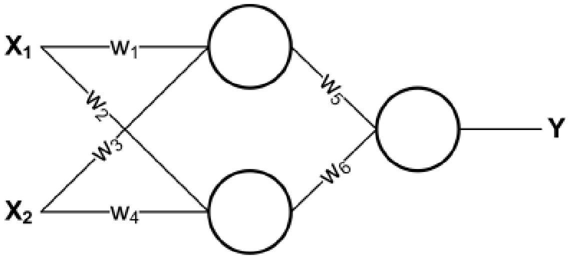
Homework is due 10/05/2018.

Problem 1 (10 points): Perceptron

Implement the linear perceptron using stochastic gradient descent (SGD) or gradient descent (GD). Download the dataset“data3.mat”. Use the whole data set as training, where each row consists of the feature vector x followed by the label $y \in \{-1, 1\}$ (last column). Show with figures the resulting linear decision boundary on the $2d$ x data. Show the evolution of binary classification error and the perceptron error with time (or number of iterations) from random initialization until convergence on a successful run (some random inits may not converge or may require many iterations). For GD, discuss the convergence behavior as you vary the step size (η).

Problem 2 (10 points): Neural network

For a given network unit U, A is the vector of activations of units that send their output to U, and W is the weight vector corresponding to these outputs. Consider the following neural network, consisting of two input units, a single hidden layer containing two units, and one output unit:



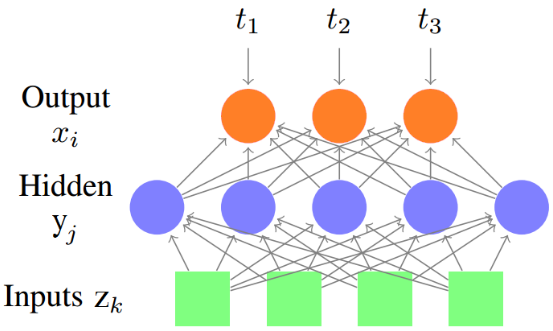
a) [4 points] Say that the network is using linear units: that is, the output of a unit is $CW^\top A$ for some fixed constant C. Let the weight values w_i be fixed. Re-design the neural network to compute the same function without using any hidden units. Express the new weights in terms of the old weights and the constant C.

b) [2 points] Is it always possible to express a neural network made up of only linear units without a hidden layer? Justify your answer.

c) [4 points] Another common activation function is a threshold, where the activation is $t(W^\top A)$, where $t(x)$ is 1 if $x > 0$ and 0 otherwise. Let the hidden units use sigmoid activation functions (activation function of U is then $(1+\exp(W^\top A))^{-1}$) and let the output unit use a threshold activation function. Find weights which cause this network to compute the XOR of X_1 and X_2 for binary-valued X_1 and X_2. Keep in mind that there is no bias term for these units.

Problem 3 (15 points): Backpropagation

Consider the following network, where x denotes output units, y denotes hidden units, and z denotes input units.



Consider:

a) [8 points] The cross-entropy error for a single example is:

$$E = -\sum_i (t_i \log(x_i) + (1-t_i)\log(1-x_i)),$$

where t is the target, and the logistic activation function for the output units is:

$$x_i = \frac{1}{1+e^{-s_i}}, \quad \text{where} \quad s_i = \sum_j y_j w_{ji},$$

where w_{ji} denotes the weight of the edge between the j^{th} hidden unit and the i^{th} output unit. Assume hidden layers also use logistic activation function.

b) [7 points] The modified cross-entopy error for a single example is:

$$E = -\sum_i t_i \log(x_i)$$

and a softmax activation function for the output units is:

$$x_i = \frac{e^{s_i}}{\sum_{c=1}^{m} e^{s_c}},$$

where m is the number of outputs and the summation is taken over all outputs. Assume hidden layers use logistic activation function.

Derive the backpropagation updates in both cases (use the above error functions defined with respect to a single training example for your derivations rather than the sum of errors over the entire training dataset).

Problem 4 (15 points): VC dimension

What is the VC dimension of the hypothesis space consisting of triangles in the 2D plane (justify your answer)? Points inside the triangle are classified as positive examples.