Machine Learning

V. Adamchik CSCI 567 Spring 2019

Discussion 5 University of Southern California

Multiclass Classification

Neural Nets

Problem 1

For a fixed multiclass problem, which of the following multiclass-to-binary reductions has the smallest testing time complexity?

(A) One-versus-all

(B) One-versus-one

(C) Tree reduction

(D) Both (A) and (C)

Problem 2

Show that one-versus-all can be seen as a special case of error-correcting-output-code (ECOC). Specifically, write down the code matrix M for ECOC for a problem with C labels so that executing ECOC is the same as doing one-versus-all. (Note: the entry of M should be either –1 or +1.)

Problem 3

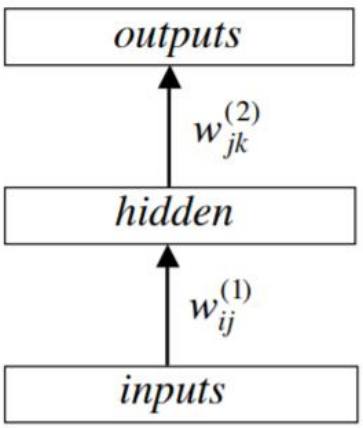
In the lecture we derive the multiclass logistic regression by minimizing the multiclass logistic loss. In this problem you need to derive the multiclass perceptron algorithm in a similar way.

(1) Define the multiclass perceptron loss.

(2) Based on (1), write down the multiclass perceptron algorithm.

Problem 4

Explain why the activation function in NN should <u>not be</u> linear.



$$out_k^{(2)} = f_2\left(\sum_j out_j^{(1)} w_{jk}^{(2)}\right)$$

$$out_j^{(1)} = f_1\left(\sum_i out_i^{(0)} w_{ij}^{(1)}\right)$$

$$out_i^{(0)} = in_i$$

f_1 and f_2 are activation functions.

Problem 5

Explain why the cost function of neural network is neither convex nor concave with respect to weights **W**?

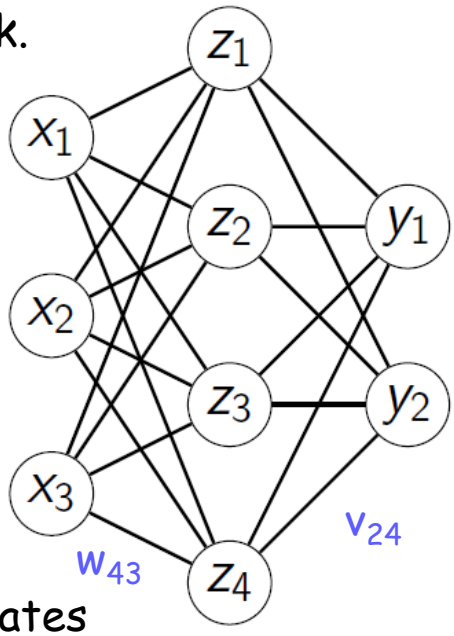
Problem 6

Consider the following neural network.

$$z_k = \tanh\left(\sum_{i=1}^{3} w_{ki} x_i\right)$$

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

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



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