Homework 2

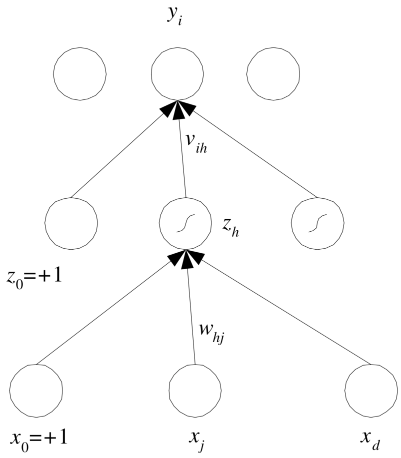
Due: Nov 11, 2020

1. (2 points)

You are given a tiny dataset as shown in the figure below. There are two categories of data labeled as y=-1 (denoted with cycles) and y=+1 (denoted with diamonds). You are required to use SVM to train a binary classifier. Please explain your solution with both theoretical derivations and optimization algorithms. (Hint: recall key concepts of SVM such as the objective function, dual problem, kernel and the SMO optimization algorithm. How to put them together to solve your problem).

2. (3 points)

Given a dataset $\chi = \{x^{(\ell)}, r^{(\ell)}\}_{\ell=1}^{N}$, we want to train the following multi-layer feedforward network for multi-class classification:



where

$$a_h^{(\ell)} = \sum_{j=0}^{d} W_{hj} x_j^{(\ell)}$$

$$z_h^{(\ell)} = \tanh(a_h^{(\ell)})$$

$$o_i^{(\ell)} = \sum_{h=0}^{H} V_{ih} z_h^{(\ell)}$$

Suppose the activation function of each hidden unit is the hyperbolic tangent (切线) function:

$$\tanh(a) = \frac{e^{a} - e^{-a}}{e^{a} + e^{-a}} ,$$

and the softmax function is applied to the outputs so that y_i approximates the posterior probability P (C_i | x):

$$y_i^{(\ell)} = \frac{\exp(o_i^{(\ell)})}{\sum_k \exp(o_k^{(\ell)})}$$

We train the neural networks by minimizing the cross-entropy loss:

$$L(\boldsymbol{W}, \boldsymbol{V} | \chi) = -\sum_{\ell=1}^{N} \sum_{k=1}^{K} r_k^{(\ell)} \log y_k^{(\ell)}$$

We use the simple gradient descent for learning the network weights.

(a) Derive the weight update rule for the second-layer weights v_{ih}.

(b) Derive the weight update rule for the first-layer weights w_{hj}.

Hint: modify the derivation in class which uses the *sigmoid* activation function.