

Homework Assignment 4

Lecturer: Kyunghyun Cho

March 9, 2018

1. RBFN for XOR problem (15 points) (a) For the 2-dimensional XOR problem, we select the following four basis vectors:

$$\mathbf{r}^1 = [-1, -1]^\top$$

$$\mathbf{r}^2 = [1, 1]^\top$$

$$\mathbf{r}^3 = [-1, 1]^\top$$

$$\mathbf{r}^4 = [1, -1]^\top.$$

Show that the XOR problem is solved by the radial basis function network with the following weight vector:

$$\mathbf{w} = [1, 1, -1, -1, 0]^\top.$$

2. Nearest-neighbour classifier by RBFN (15 points) A nearest-neighbour classifier can be constructed as a radial basis function network by selecting all the input vectors in a training set as basis vectors. In a multi-class classification setting (i.e., there are more than two categories), provide a description on how a weight matrix could be built.

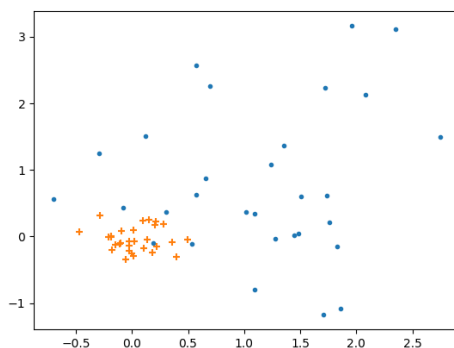
3. Adaptive RBFN (15 points) Unlike a fixed basis function network, an adaptive basis function network adapts basis vectors so as to maximize the classification accuracy (i.e. to minimize the empirical cost.) In order to do so, we need to be able to compute the gradient of the (logistic regression) distance function with respect to each and every basis vector. Derive this gradient

$$\nabla_{\mathbf{r}^k} D(y^*, M, \phi(\mathbf{x})),$$

assuming that M is a logistic regression classifier and that

$$\phi(\mathbf{x}) = \begin{bmatrix} \exp(-(\mathbf{x} - \mathbf{r}^1)^2) \\ \vdots \\ \exp(-(\mathbf{x} - \mathbf{r}^K)^2) \end{bmatrix}.$$

4. Adaptive RBFN on toy data (15 points) Suppose that we are fitting an Adaptive RBFN with the formulation in Problem 3 on the following toy dataset whose positive and negative samples are generated from two Gaussian distributions with very different variances:



Would the Adaptive RBFN work (i.e. propose a reasonable decision boundary)? If so, how would the parameters (i.e. bases, weights, and biases) possibly look like? If not, how can we improve it? *You don't have to answer with math equations; plain English is fine.*

5. k-NN classifier with Scikit-learn and RBFN with Stochastic Gradient Descent (20 points each)