Practice Set 3

1. Consider a one-dimensional two-category classification problem with equal priors. Three i.i.d training observations were collected: D_1 ={1, 2, 6} and D_2 = {4, 5, 7} for ω_1 and ω_2, respectively. It is desired to classify a test pattern using the Parzen Window technique. Recall that in this approach, the estimate for $\mathbf{x}$ over the set of training samples $\{\mathbf{x}_i\}_{i=1}^{n}$ is given by $p_n(\mathbf{x})= \frac{1}{n}\sum_{i=1}^{n}\phi(\mathbf{x}-\mathbf{x}_i)$.

   (i) Using the window function $\phi(x) = \frac{1}{2}\exp^{-|x|}$, classify the pattern $x = 4.5$. You may assume the scenario of zero-loss loss function.

   (ii) Verify whether the function $\phi(x)$ is a probability density function.

2. You are given 2 training examples (comprising 2 dimensions) for classes ω_1 and ω_2 .

$$\omega_1 : \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \omega_2 : \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

   Compute the optimal weight vector $\mathbf{w}$ (projection line in a reduced dimension) using the Fisher's criterion.

   The inverse of 2×2 matrix: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

3. You are given a data set $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$ of size N. Each input $\mathbf{x}_i$ is d-dimensional and its corresponding target y_i takes one of the 2 values (+1 or -1). A modified version of the SVM classifier is obtained by reformulating the minimization problem of the traditional problem as:

$$\min_{\mathbf{w},b,\xi,\rho} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{1}{N}\sum_{i=1}^{N}\xi_i - \rho$$

   subject to the constraints:

$$y_i(\mathbf{w}^T\mathbf{x}_i + b] \geq \rho - \xi_i \qquad i = 1, 2, 3.....N$$

$$\xi_i \geq 0 \qquad i = 1, 2, 3.....N$$

$$\rho \geq 0$$

   (a) Write down the Lagrangian Function $L(\mathbf{w}, b, \boldsymbol{\xi}, \rho)$ corresponding to this optimization problem. Here $\boldsymbol{\xi} = \{\xi_1, \xi_2, ...\xi_N\}$.

   (b) State the KKT conditions to the primal problem, that are to be satisfied at optimality.

4. Consider a data set $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$, where each input $\mathbf{x}_i$ is d-dimensional and its corresponding target y_i is an integer valued scalar, taking values +1 or -1. Additionally, assume that each data pair $(\mathbf{x}_i, y_i)$ is associated with a **known** weight r_i, (where $r_i > 0$). Accordingly, the sum of squares error function can be written as $E(\mathbf{w}) = \sum_{i=1}^{N} r_i(y_i - \mathbf{w}^T\mathbf{x}_i)^2$. An iterative approach to this minimization/learning problem is the method of gradient descent. Using this technique, write an equation that updates the weight vector $\mathbf{w}(j-1)$ in the j^{th} iteration. You may assume the 'batch-mode' learning paradigm for updating the weights.