Assignment 2 - Problem 1

Deriving formulas for Stochastic Gradient-Based method for training an RBF NN.

Given Error Cost function

$$J(n) = \frac{1}{2} |e(n)|^2 = \frac{1}{2} [y_d(n) - y(n)]^2 \rightarrow ①$$

$y_d(n)$ = desired output

$$y(n) = \sum_{K=1}^{N} w_K(n) \cdot \phi\{x(n), c_K, \sigma_K\} \rightarrow \text{Actual output}$$

RBF seed function $\phi\{x(n), c_K, \sigma_K\}$ is chosen to be Gaussian kernel, hence

$$y(n) = \sum_{K=1}^{N} w_K(n) \, e^{-\frac{\|x(n) - c_K(n)\|^2}{2\sigma_K^2(n)}} \rightarrow ②$$

where,

$c_K(n)$ is the center vector for K^{th} radial function

$\sigma_K(n)$ is the spread parameter

Updated parameters of the network are given by the following equations

$$w(n+1) = w(n) - \mu_w \frac{\partial J(n)}{\partial w}\Big|_{w = w(n)} \rightarrow Ⓐ$$

$$c_k(n+1) = c_k(n) - \mu_c \frac{\partial}{\partial c_k} J(n) \Big|_{c_k = c_k(n)} \rightarrow (B)$$

$$\sigma_k(n+1) = \sigma_k(n) - \mu_\sigma \frac{\partial}{\partial \sigma_k} J(n) \Big|_{\sigma_k = \sigma_k(n)} \rightarrow (C)$$

μ_w, μ_c, μ_σ are appropriate learning parameters.

$\Rightarrow$ Considering the chain rule of differentiation,

$$\frac{\partial J(n)}{\partial w} = \frac{\partial J(n)}{\partial e(n)} \cdot \frac{\partial e(n)}{\partial y(n)} \cdot \frac{\partial y(n)}{\partial w}$$

From (1),
$$\frac{\partial J(n)}{\partial e(n)} = \frac{\partial}{\partial e(n)} \left[\frac{1}{2} | e(n) |^2 \right] = \frac{1}{2} \cdot 2 \cdot e(n)$$

$$= e(n) \rightarrow (3)$$

From (2),
$$\frac{\partial y(n)}{\partial w} = \frac{\partial}{\partial w} \left[\sum_{K=1}^{N} w_k(n) \, e^{\frac{-|x(n) - c_k(n)|^2}{2\sigma_k^2(n)}} \right]$$

$$= \sum_{K=1}^{N} e^{\frac{-|x(n) - c_k(n)|^2}{2\sigma_k^2(n)}} \rightarrow (4)$$

From (1),
$$\frac{\partial e(n)}{\partial y(n)} = \frac{\partial}{\partial y(n)} \left[y_d(n) - y(n) \right] = -1 \rightarrow (5)$$

From (3), (4) and (5)

$$\frac{\partial J(n)}{\partial w} = e(n) \cdot (-1) \cdot \sum_{K=1}^{N} \phi \{ x(n), c_k, \sigma_k \}$$

From (A)

$$w(n+1) = w(n) - \mu_w \cdot \left[-e(n) \sum_{K=1}^{N} \phi \{ x(n), c_k, \sigma_k \} \right]$$

$$w(n+1) = w(n) + \mu_w \, e(n) \, \varphi(n) \rightarrow (6)$$

where $\varphi(n) = \left[\phi\{x(n), c_1, \sigma_1\}, \dots \phi\{x(n), c_N, \sigma_N\} \right]^T$

$$\Rightarrow \frac{\partial J(n)}{\partial c_K} = \frac{\partial J(n)}{\partial e(n)} \cdot \frac{\partial e(n)}{\partial y(n)} \cdot \frac{\partial y(n)}{\partial c_K}$$

From ②, $$\frac{\partial y(n)}{\partial c_K} = \frac{\partial}{\partial c_K} \left[\sum_{K=1}^{N} w_K(n) \cdot e^{-\frac{|x(n) - c_K(n)|^2}{2\sigma_K^2(n)}} \right]$$

$$= \sum_{K=1}^{N} w_K(n) \cdot \underbrace{e^{-\frac{|x(n) - c_K(n)|^2}{2\sigma_K^2(n)}}}_{\phi\{x(n), c_K, \sigma_K\}} \cdot \left[\frac{-2\,|x(n) - c_K(n)|}{2\sigma_K^2(n)} \right] \cdot (-1)$$

Re-writing,

$$\frac{\partial y(n)}{\partial c_K} = \frac{w_K(n)}{\sigma_K^2(n)} \phi\{x(n), c_K, \sigma_K\} \cdot |x(n) - c_K(n)|$$

From ⑧, $$c_K(n+1) = c_K(n) - \mu_c\,(-1)\, e(n) \cdot \frac{\partial y(n)}{\partial c_K}$$

$$c_K(n+1) = c_K(n) + \mu_c \frac{e(n)\, w_K(n)}{\sigma_K^2(n)} \phi\{x(n), c_K, \sigma_K\} \underbrace{[x(n) - c_K(n)]}_{⑦}$$

$$\Rightarrow \frac{\partial J(n)}{\partial \sigma_K} = \frac{\partial J(n)}{\partial e(n)} \cdot \frac{\partial e(n)}{\partial y(n)} \cdot \frac{\partial y(n)}{\partial \sigma_K}$$

From ②, $$\frac{\partial y(n)}{\partial \sigma_K} = \frac{\partial}{\partial \sigma_K} \left[\sum_{K=1}^{N} w_K(n) \cdot e^{-\frac{|x(n) - c_K(n)|^2}{2\sigma_K^2(n)}} \right]$$

$$= \sum_{K=1}^{N} w_K(n) \cdot \underbrace{e^{-\frac{|x(n) - c_K(n)|^2}{2\sigma_K^2(n)}}}_{\phi\{x(n), c_K, \sigma_K\}} \cdot \left[-|x(n) - c_K(n)|^2 \right] \cdot \left[\frac{-2}{2\sigma_K^3(n)} \right]$$

$$\frac{\partial y(n)}{\partial \sigma_K} = \frac{w_K(n)}{\sigma_K^3(n)} \phi\{x(n), C_K, \sigma_K\} \|x(n) - C_K(n)\|^2$$

From ⑥,

$$\sigma_K(n+1) = \sigma_K(n) - \mu_\sigma (-1)(e(n)) \cdot \frac{\partial y(n)}{\partial \sigma_K}$$

$$\sigma_K(n+1) = \sigma_K(n) + \mu_\sigma \frac{e(n) w_K(n)}{\sigma_K^3(n)} \phi\{x(n), C_K(n), \sigma_K\} \cdot \|x(n) - C_K(n)\|^2 \rightarrow ⑧$$

from ⑥, ⑦, ⑧

$$w(n+1) = w(n) + \mu_w e(n) \psi(n)$$

$$C_K(n+1) = C_K(n) + \mu_c \frac{e(n) w_K(n)}{\sigma_K^2(n)} \phi\{x(n), C_K(n), \sigma_K\} [x(n) - C_K(n)]$$

$$\sigma_K(n+1) = \sigma_K(n) + \mu_\sigma \frac{e(n) w_K(n)}{\sigma_K^3(n)} \phi\{x(n), C_K(n), \sigma_K\} \|x(n) - C_K(n)\|^2$$

where,

$$\psi(n) = [\phi\{x(n), C_1, \sigma_1\}, \ldots \ldots \phi\{x(n), C_N, \sigma_N\}]^T$$