Tik-61.261 Principles of Neural Computing

Raivio, Venna

Exercise 4

1. Let the error function be
$$\mathcal{E}(\mathbf{w}) = w_1^2 + 10w_2^2,$$
where w_1 and w_2 are the components of the two-dimensional parameter vector $\mathbf{w}$. Find the minimum value of $\mathcal{E}(\mathbf{w})$ by applying the steepest descent method. Use $\mathbf{w}(0) = [1, 1]^T$ as an initial value for the parameter vector and the following constant values for the learning rate:

   (a) $\alpha = 0.04$
   (b) $\alpha = 0.1$
   (c) $\alpha = 0.2$
   (d) What is the condition for the convergence of this method?

2. Show that the application of the Gauss-Newton method to the error function
$$\mathcal{E}(\mathbf{w}) = \frac{1}{2}\left[\delta\|\mathbf{w} - \mathbf{w}(n)\|^2 + \sum_{i=1}^{n} e_i^2(\mathbf{w})\right]$$
yields the the following update rule for the weights:
$$\Delta\mathbf{w} = -\left[\mathbf{J}^T(\mathbf{w})\mathbf{J}(\mathbf{w}) + \delta\mathbf{I}\right]^{-1}\mathbf{J}^T(\mathbf{w})\mathbf{e}(\mathbf{w}).$$
All quantities are evaluated at iteration step n. (Haykin 3.3)

3. The normalized LMS algorithm is described by the following recursion for the weight vector:
$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \frac{\eta e(n)\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2},$$
where η is a positive constant and $\|\mathbf{x}(n)\|$ is the Euclidean norm of the input vector $\mathbf{x}(n)$. The error signal $e(n)$ is defined by
$$e(n) = d(n) - \hat{\mathbf{w}}(n)^T\mathbf{x}(n),$$
where $d(n)$ is the desired response. For the normalized LMS algorithm to be convergent in the mean square, show that $0 < \eta < 2$. (Haykin 3.5)

4. The ensemble-averaged counterpart to the sum of error squares viewed as a cost function is the mean-square value of the error signal:
$$J(\mathbf{w}) = \frac{1}{2}E[e^2(n)] = \frac{1}{2}E[(d(n) - \mathbf{x}^T(n)\mathbf{w})^2].$$

   (a) Assuming that the input vector $\mathbf{x}(n)$ and desired response $d(n)$ are drawn from a stationary environment, show that
$$J(\mathbf{w}) = \frac{1}{2}\sigma_d^2 - \mathbf{r}_{\mathbf{x}d}^T\mathbf{w} + \frac{1}{2}\mathbf{w}^T\mathbf{R}_{\mathbf{x}}\mathbf{w},$$
where $\sigma_d^2 = E[d^2(n)]$, $\mathbf{r}_{\mathbf{x}d} = E[\mathbf{x}(n)d(n)]$, and $\mathbf{R}_{\mathbf{x}} = E[\mathbf{x}(n)\mathbf{x}^T(n)]$.

   (b) For this cost function, show that the gradient vector and Hessian matrix of $J(\mathbf{w})$ are as follows, respectively:
$$\mathbf{g} = -\mathbf{r}_{\mathbf{x}d} + \mathbf{R}_{\mathbf{x}}\mathbf{w} \quad \text{and}$$
$$\mathbf{H} = \mathbf{R}_{\mathbf{x}}.$$

(c) In the LMS/Newton algorithm, the gradient vector $\mathbf{g}$ is replaced by its instantaneous value. Show that this algorithm, incorporating a learning rate parameter η, is described by

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \eta \mathbf{R}_{\mathbf{x}}^{-1} \mathbf{x}(n) \left[d(n) - \mathbf{x}^T(n) \hat{\mathbf{w}}(n) \right].$$

The inverse of the correlation matrix $\mathbf{R}_{\mathbf{x}}$, assumed to be positive definite, is calculated ahead of time. (Haykin 3.8)

5. A linear classifier separates n-dimensional space into two classes using a $(n-1)$-dimensional hyperplane. Points are classified into two classes, ω_1 or ω_2, depending on which side of the hyperplane they are located.

(a) Construct a linear classifier which is able to separate the following two-dimensional samples correctly:

$$\begin{aligned} \omega_1 &: \{[2,1]^T\}, \\ \omega_2 &: \{[0,1]^T, [-1,1]^T\}. \end{aligned}$$

(b) Is it possible to construct a linear classifier which is able to separate the following samples correctly?

$$\begin{aligned} \omega_1 &: \{[2,1]^T, [3,2]^T\}, \\ \omega_2 &: \{[3,1]^T, [2,2]^T\} \end{aligned}$$

Justify your answer.