MEAN SQUARED ERROR (OR LEAST SQUARES) ALGORITHMS

MSE Estimation

- Given a vector $\mathbf{x}$, the classifier yields the output $\mathbf{w}^T\mathbf{x}$
  - A threshold can be accommodated by the vector extension
- The desired output is $y(\mathbf{x}) = \pm 1$

MSE Estimation

- Problem:

$$\text{Minimize } J(\mathbf{w}) = E\left[\left(y - \mathbf{w}^T\mathbf{x}\right)^2\right]$$

- Solution

$$\hat{\mathbf{w}} = R_{\mathbf{x}}^{-1} \times E[\mathbf{x}y]$$

Here, the autocorrelation matrix and the cross-correlation vector are

$$R_{\mathbf{x}} = E\left[\mathbf{x}\mathbf{x}^T\right] = \begin{bmatrix} E[x_1x_1] & \cdots & E[x_1x_l] \\ E[x_2x_1] & \cdots & E[x_2x_l] \\ \vdots & \vdots & \vdots \\ E[x_lx_1] & \ldots & E[x_lx_l] \end{bmatrix} \text{ and } E[\mathbf{x}y] = \begin{bmatrix} E[x_1y] \\ \vdots \\ E[x_ly] \end{bmatrix}$$

MSE Estimation

- Multiclass generalization
  - Design M linear discriminant functions $g_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x}$ according to the MSE criterion
    - The output responses are chosen so that $y_i = 1$ if $\mathbf{x} \in \omega_i$ and $y_i = 0$ otherwise.
    - If $M = 2$, it provides the decision hyperplane $\mathbf{w}^T\mathbf{x} \equiv (\mathbf{w}_1 - \mathbf{w}_2)^T\mathbf{x}$, which tries to yield ± 1 according to the class of $\mathbf{x}$
  - Classify an input vector $\mathbf{x}$ into ω_i that yields the highest response $g_i(\mathbf{x})$

MSE Estimation

- Multiclass generalization
  - The MSE criterion
    - Let $\mathbf{y}^T = [y_1, \dots, y_M]$ and $W = [\mathbf{w}_1, \dots, \mathbf{w}_M]$
    - Then, we have

$$\hat{W} = \arg\min_W E\left[\left\|\mathbf{y} - W^T\mathbf{x}\right\|^2\right]$$

$$= \arg\min_W E\left[\sum_{i=1}^{M}\left(y_i - \mathbf{w}_i^T\mathbf{x}\right)^2\right]$$

Method of Least Squares

- Given training samples $(\mathbf{x}_i, y_i)$, minimize the least squares criterion

$$J(\mathbf{w}) = \sum_{i=1}^{N} \left(y_i - \mathbf{x}_i^T \mathbf{w}\right)^2$$

- Solution

$$\hat{\mathbf{w}} = \left(X^T X\right)^{-1} X^T \mathbf{y}$$

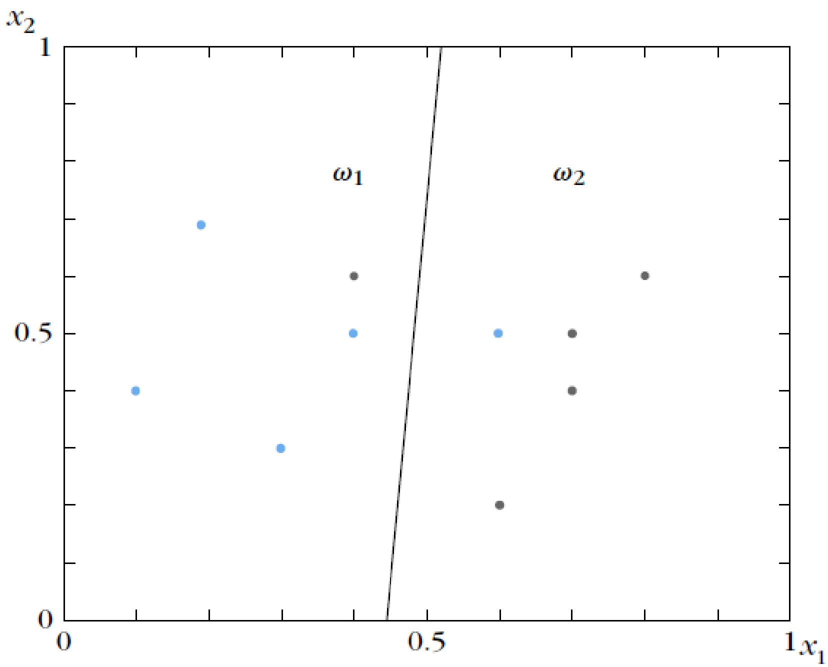
where

$$X = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{11} & \cdots & x_{1l} \\ x_{21} & x_{22} & \cdots & x_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nl} \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Method of Least Squares

Example 3.4

Class ω_1 consists of the two-dimensional vectors $[0.2, 0.7]^T$, $[0.3, 0.3]^T$, $[0.4, 0.5]^T$, $[0.6, 0.5]^T$, $[0.1, 0.4]^T$ and class ω_2 of $[0.4, 0.6]^T$, $[0.6, 0.2]^T$, $[0.7, 0.4]^T$, $[0.8, 0.6]^T$, $[0.7, 0.5]^T$. Design the sum of error squares optimal linear classifier $w_1x_1 + w_2x_2 + w_0 = 0$.



MSE Regression (nonlinear)

- Regression task
  - A problem of designing a function $g(\mathbf{x})$, based on a set of training data points $(y_i, \mathbf{x}_i)$, so that the predicted value
  $$\hat{y} = g(\mathbf{x})$$
  is as close to the true value y as possible
  - y need not be ± 1

- MSE regression
  $$\hat{y} = \arg\min_{\tilde{\mathrm{y}}} E[\|y - \tilde{y}\|^2] = E[y|\mathbf{x}]$$
  - The MSE regression function is linear, when $(y, \mathbf{x})$ are jointly Gaussian