Chapter-5 : Single-layer Networks: Classification

5.1 Discriminant Functions

A discriminant function takes input x and assigns it to one of K classes, denoted C_K.

5.1.1 Two Classes

$$y(x) = w^T x + w_0$$

↳ input x is assigned to C_1 if $y(x) \geq 0$ and to C_2 otherwise.

The decision boundary is defined where $y(x) = 0$

Let x_A & x_B be on the boundary then.

$$y(x_A) = y(x_B) = 0$$

$$\Rightarrow \quad y(x_A) - y(x_B) = 0$$

$$(w^T x_A + w_0) - (w^T x_B + w_0) = 0$$

$$w^T (x_A - x_B) = 0$$

Note that $x_A - x_B$ also lies on the boundary

This tells us the w is orthogonal to any vector lying in the plane.

$$x_\perp = r_0 \underbrace{\frac{w}{\|w\|}}_{}$$

→ unit vec in direction of w.

$$w^T\left(r_0 \frac{w}{\|w\|}\right) + w_0 = 0$$

$$r_0 \|w\| + w_0 = 0$$

$$r_0 = \frac{-w_0}{\|w\|}$$

5.1.2 Multiple classes

We have K classes $C_1, \dots, C_K$

$$y_k(x) = w_k^T x + w_{k0} \quad (k = 1, \dots, K)$$

$$\hat{k}(x) = \underset{k}{\operatorname{argmax}}\, y_k(x)$$

The boundary b/we C_i & C_j is where the model is indifferent:

$$y_i(x) = y_j(x) \iff (w_i - w_j)^T x + (w_{i0} + w_{j0}) = 0$$

A set $S \subseteq \mathbb{R}^D$ is convex if $\forall\, x_A, x_B \in S$:

$$\lambda x_A + (1-\lambda) x_B \in S \quad , \quad \forall \lambda \in [0,1]$$

$$y_K(x) = w_K^T x + w_{K0}$$

Region belonging to class C_K is:

$$R_K = \{ x \mid y_K(x) > y_j(x) \;,\; \forall j \neq k \}$$

Let $x_A, x_B \in R_K$

We define the point on the line joining them as:

$$\hat{x} = \lambda x_A + (1-\lambda) x_B \quad , \quad \lambda \in [0,1]$$

$$\begin{aligned}
y_K(\hat{x}) &= w_K^T \hat{x} + w_{K0} \\
&= w_K^T \left(\lambda x_A + (1-\lambda) x_B \right) + w_{K0} \\
&= \lambda \left(w_K^T x_A + w_{K0} \right) + (1-\lambda) \left(w_K^T x_B + w_{K0} \right) \\
&= \lambda\, y_K(x_A) + (1-\lambda)\, y_K(x_B)
\end{aligned}$$

5.1.4 Least Squares for Classification

vector-matrix notation

$$\tilde{x} = \begin{bmatrix} 1 \\ x \end{bmatrix}, \quad \tilde{w}_k = \begin{bmatrix} w_{k0} \\ w_k \end{bmatrix}$$

$$\tilde{W} = \begin{bmatrix} w_{10} & w_{20} & \cdots\cdots & w_{k0} \\ w_1 & w_2 & & w_k \end{bmatrix}$$

$$\tilde{W} = [\tilde{w}_1, \tilde{w}_2, \ldots\ldots, \tilde{w}_K]$$

$$\boxed{y(x) = \tilde{W}^T \tilde{x}}$$

We have N training samples $\{(x_n, t_n)\}_{n=1}^{N}$
where each $t_n = (0, 0, \ldots, 1, \ldots, 0)^T$

- $\tilde{X} = N \times (D+1) \longrightarrow$ input rows $\tilde{x}_n^T$
- $T = N \times K \longrightarrow$ target rows t_n^T

$$E_D(\tilde{W}) = \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} (y_{n_k} - t_{n_k})^2$$