binary classification: $w \longrightarrow f(x) = h(x) = y(x) = w^T x > 0$?

Yes $\to +1$; No $\to -1$.

multiclass classification: $K \geq 2$ classes.

① <u>One-vs-rest</u>: train K one-vs-rest binary classifiers

- for a class k, <u>train parameters</u> $w^{(k)}$ to classify an arbitrary example as to whether it belongs to class k ($+1$) or not (-1).

  K sets of parameters $w^{(k)}$, $k = 1..K$.

  examples labelled as k are $\oplus$ve, everything else is $\ominus$ve.

- given test example x, $\hat{y}(x) = \operatorname{argmax}_{k=1..K} \, {w^{(k)}}^T x$

② <u>One-vs.-one</u>: train $\binom{K}{2}$ classifiers.

- given test example x, run it through all $\binom{K}{2}$ classifiers, output the "<u>class with most wins</u>".

③ Compute a score $s(x,k) = w^T \phi(x,k)$.

- train w
- test : $\hat{y}(x) = \underset{k=1..K}{\operatorname{argmax}}\ s(x,k)$.

<u>Multiclass</u> : $K \geq 2$ classes, X belongs to only 1 class

<u>Multilabel</u> : $K \geq 2$ classes, X may belong to 1 or more classes.

decision boundary $\{x \mid \overbrace{w^T x + b}^{y(x)} = 0\} = H$

distance from x to $H = \dfrac{y(x)}{\|w\|}$

the margin for the classifier (w, b) is $\min\limits_{n=1..N} \dfrac{|y(x_n)|}{\|w\|}$

$$\boxed{\text{margin}(w, b) = \min_{n=1..N} \frac{|w^T x_n + b|}{\|w\|}}$$

The max-margin classifier $(w^*, b^*) = \arg\max\limits_{w,b} \text{margin}(w, b)$

$$\boxed{(w^*, b^*) = \arg\max_{w,b} \min_{n=1..N} \frac{|w^T x_n + b|}{\|w\|}}$$

$\text{margin}(w, b) = \min\limits_{n=1..N} \dfrac{|w^T x_n + b|}{\|w\|}$;

$\text{margin}(cw, cb) = \min\limits_{n=1...N} \dfrac{|c w^T x_n + cb|}{\|cw\|} = \dfrac{\cancel{c}|w^T x_n + b|}{\cancel{c}\|w\|}$

same!

Let's (wlg) consider that $|w^T x_n + b| = 1$, for the closest sample x_n.

$$|w^T x_n + b| \geq 1, \quad \forall n = 1 \dots N$$

max-margin

Then the (min - max) opt. problem becomes:

$$\boxed{\begin{aligned} (w^*, b^*) &= \underset{w,b}{\arg\max} \frac{1}{\|w\|} \\ \text{subj. to: } & |w^T x_n + b| \geq 1, \quad \forall n = 1 \dots N \end{aligned}}$$

use assumption of separable case.

$$\Downarrow$$

$$\boxed{\begin{aligned} (w^*, b^*) &= \underset{w,b}{\arg\min} \frac{1}{2}\|w\|^2 \\ \text{subj. to: } & t_n (w^T x_n + b) \geq 1 \end{aligned}}$$

SVM optimization prob.

$$\min_{\underbrace{w,b}_{x}} J(w,b) = \underbrace{\frac{1}{2}\|w\|^2}_{} \longrightarrow f_0(x)$$

subj. to $\underbrace{1 - t_n(w^T x_n + b)}_{f_n(w,b)} \leq 0, \quad \forall n = 1..N$

Dual params. $\alpha_n, \quad n = 1..N. \quad \alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]$

Define Lagrangian $\mathcal{L}(\underbrace{w,b},\alpha) = \frac{1}{2}\|w\|^2 + \sum_{n=1}^{N} \alpha_n \left(1 - t_n(w^T x_n + b)\right)$

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \Rightarrow w + \sum_{n=1}^{N} -\alpha_n t_n x_n = 0 \Rightarrow \boxed{w = \sum_{n=1}^{N} \alpha_n t_n x_n}$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow \sum_{n=1}^{N} \alpha_n \cdot (-t_n) = 0 \Rightarrow \sum_{n=1}^{N} \alpha_n t_n = 0$$

like in Kernel Perceptron!