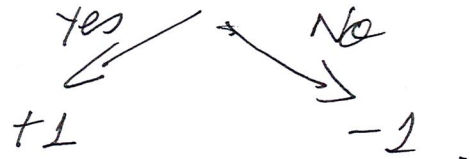


binary classification: $w \rightarrow f(x) = h(x) = y(x) = w^T x > 0?$



multiclass classification: $K \geq 2$ classes.

① One-vs-rest: train K one-vs-rest classifiers

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

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

examples labeled as k are +1, everything else is -1.

given test example x , $\hat{y}(x) = \arg \max_k w^{(k)T} x$

② One-vs-one: train $\binom{K}{2}$ classifiers.

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

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

· train w

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

Multiclass : $K \geq 2$ classes, x belongs to only 1 class

Multilabel : $K \geq 2$ 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 $H = \frac{y(x)}{\|w\|}$

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

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

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

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

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

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

same!

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

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

Then the ^{max-margin} (min-max) opt. problem becomes:

$$(w^*, b^*) = \arg \max_{w, b} \frac{1}{\|w\|}$$

$$\text{subj. to: } |w^T x_n + b| \geq 1, \quad \forall n = 1 \dots N$$



$$(w^*, b^*) = \arg \min_{w, b} \frac{1}{2} \|w\|^2$$

$$\text{subj. to: } t_n (w^T x_n + b) \geq 1$$

use assumption of separable case.

SVM optimization prob.

$$\min_{\underline{x}} f(\underline{w}, b) = \frac{1}{2} \|\underline{w}\|^2 \rightarrow f_0(\underline{x})$$

$$\text{subject to } \underbrace{1 - t_n (\underline{w}^T \underline{x}_n + b)}_{f_n(\underline{w}, b)} \leq 0, \quad \forall n = 1 \dots N$$

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

$$\text{Define Lagrangian } \mathcal{L}(\underline{w}, b, \alpha) = \frac{1}{2} \|\underline{w}\|^2 + \sum_{n=1}^N \alpha_n (1 - t_n (\underline{w}^T \underline{x}_n + b))$$

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

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

like in Kernel Perceptron!