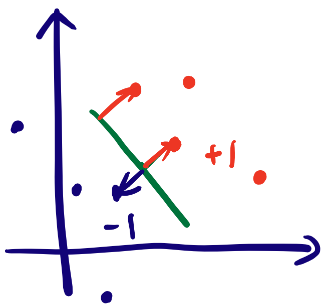
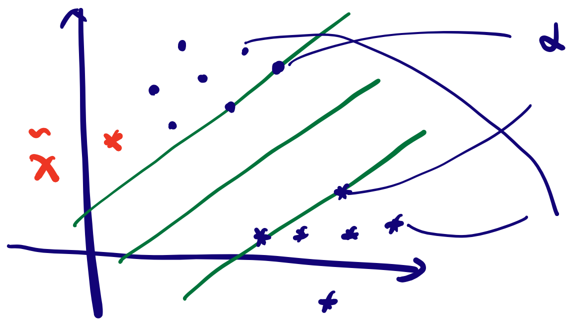




$\alpha_i \neq 0, \; y_i(w^T x_i + b) = 1$

Support vectors

$\alpha_i = 0$

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

Only a few non-zero α_i

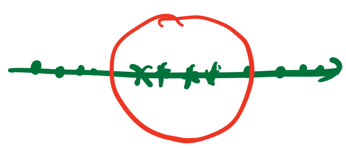
$\tilde{x}$: new testing sample

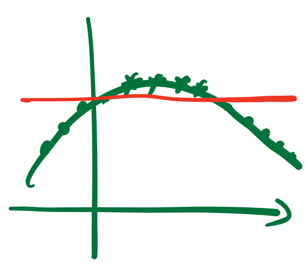
$$w^T \tilde{x} + b = \sum_{i=1}^{N} \alpha_i y_i x_i \tilde{x} + b$$

$x_i \tilde{x} \Rightarrow k(x_i, \tilde{x})$ inner product

Gaussian kernel

$$k(x, \gamma) = \exp\left(-\frac{\|x-\gamma\|^2}{2\sigma^2}\right) \qquad x \Rightarrow \phi(x) = \begin{bmatrix} x \\ x^2 \\ \vdots \\ x^d \end{bmatrix}$$





non-seperable SVM

$$\min L(w, b, \xi, \alpha, r)$$

$$= \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_n - \sum_{n=1}^{N} \alpha_n \left(y_n (w^T x_n + b) - 1 + \xi_n \right) - \sum_{n=1}^{N} r_n \xi_n$$

s.t. $\alpha_n \geq 0$, $r_n \geq 0$, $n = 1 \cdots N$

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{n=1}^{N} \alpha_n y_n x_n$$

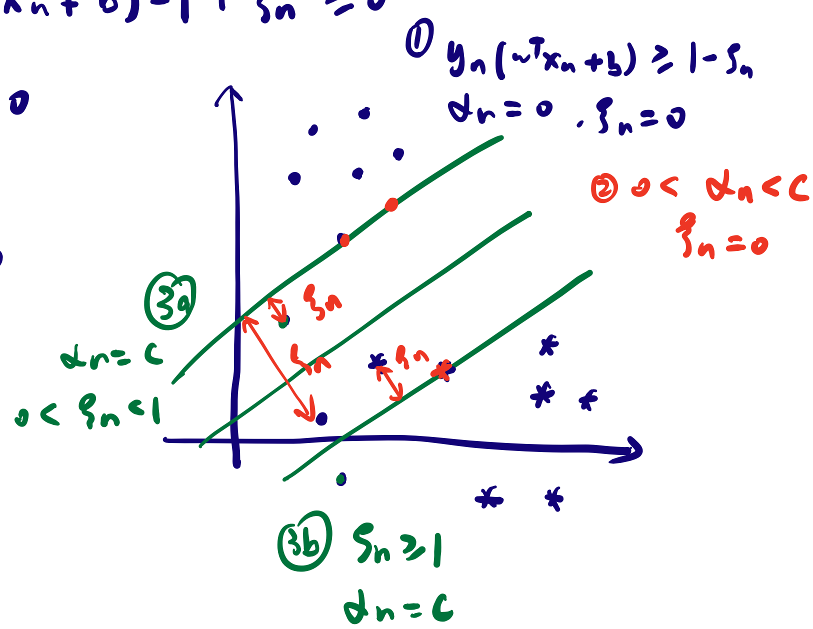
$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{n=1}^{N} \alpha_n y_n = 0$$

$$\frac{\partial L}{\partial \xi_n} = 0 \Rightarrow C - \alpha_n - r_n = 0 \Rightarrow r_n = C - \alpha_n$$

KKT conditions

$$\begin{cases} \alpha_n \left(y_n (w^T x_n + b) - 1 + \xi_n \right) = 0 \\ \alpha_n \geq 0 \\ y_n (w^T x_n + b) - 1 + \xi_n \geq 0 \end{cases}$$

$$\begin{cases} r_n \xi_n = 0 \\ r_n \geq 0 \\ \xi_n \geq 0 \end{cases}$$



① $\alpha_n = 0 \Rightarrow r_n = C - \alpha_n \Rightarrow \xi_n = 0$

$y_n(w^T x_n + b) \geq 1 - \xi_n$.

Correct samples

② $0 < \alpha_n < C \Rightarrow y_n(w^T x_n + b) = 1 - \xi_n$

$0 < C - r_n < C \Rightarrow 0 < r_n < C \Rightarrow \xi_n = 0$

$y_n(w^T x_n + b) = 1$

points on the margin

③ $y_n(w^T x + b) = 1 - \xi_n$

$\alpha_n = C \Rightarrow r_n = 0$

$\xi_n \geq 0$

ⓐ inside the margin. $\alpha_n = C$

$0 < \xi_n < 1$

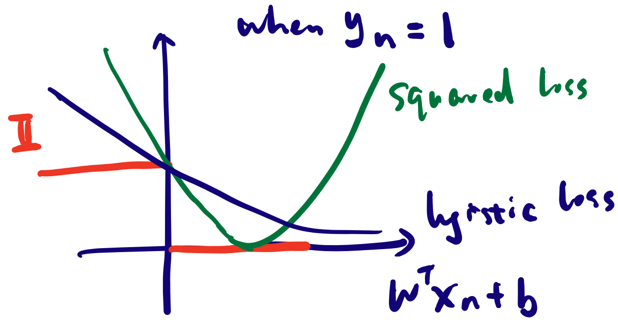
ⓑ wrong sample $\alpha_n = C$

$\xi_n \geq 1$

Linear classification

indicator $\mathbb{I}(y_n(\underline{w^T x_n + b}) < 0) = \begin{cases} 1, & \text{negative} \\ 0, & \text{positive} \end{cases}$

when $y_n = 1$



$$\min_{w,b} L(w,b) = \min \mathbb{I}\left(y_n(w^T x_n + b) < 0\right) + \lambda R(w,b)$$

- squared loss

$$\left(1 - y_n(w^T x_n + b)\right)^2 = \left(1 - (w^T x_n + b)\right)^2$$

- logistic regression

$$\log\left(1 + \exp(-(w^T x + b)\right)$$

- svm loss hinge loss

if $w^T x + b \geq 1$ $\quad \xi_n = 0$

if $w^T x + b < 1$ $\quad \xi_n = 1 - (w^T x + b)$