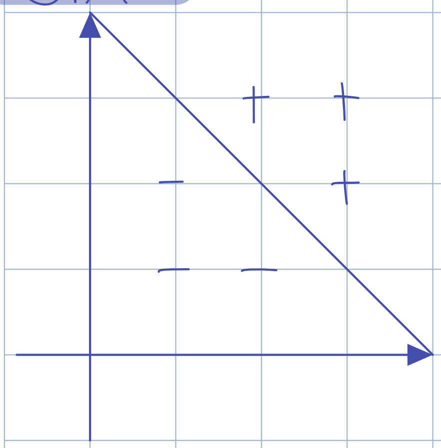
SVM



$C: w^T x + b = 0$

$$D = \{(\underline{x_1}, y_1), (\underline{x_2}, y_2), \dots (\underline{x_n}, y_n)\}$$

Primal Form

$$\min \frac{1}{2} \underline{w^T w}$$

Subject

$$y_n(\underline{w^T x_n} + b) \geq 1$$

Dual Form

$$\min_{\alpha_1, \dots, \alpha_n} \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m x_n^T x_m - \sum_{n=1}^{N} \alpha_n$$

$$\alpha_i \geq 0$$

$$\sum_{n=1}^{N} \alpha_n y_n = 0$$

$$\underline{w^*} = \sum_{n=1}^{N} \alpha_n y_n \underline{x_n}$$

Support Vectors

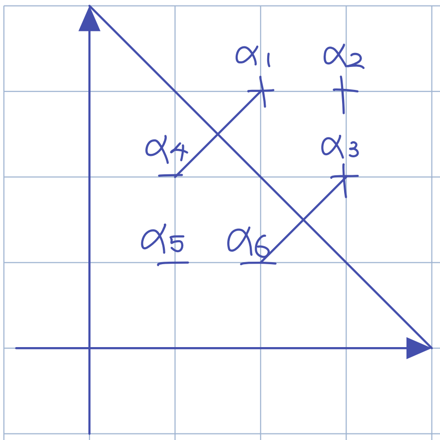
$$\underline{w^*} = \sum_{\forall n : \alpha_n > 0} \alpha_n y_n \underline{x_n}$$

KKT Conditions

for $n = 1, 2, \dots, N$

$$\alpha_n \underbrace{(y_n w^T x_n + y_n b - 1)} = 0$$

If $\alpha_n > 0 \longrightarrow y_n \underline{w^T x_n} + y_n b = 1$



(x_n, y_n) is a point with minimum distance to classification boundary

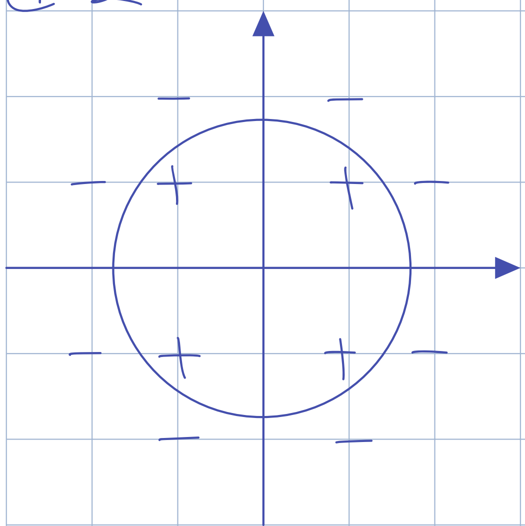
Consider any support vector $\alpha_n > 0$

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

$$b = \frac{1}{y_n}\left(1 - y_n w^T \underline{x_n}\right)$$

Kernel Trick Of SVM

$d=2$



$$X \xrightarrow[\text{transformation}]{\text{non linear}} Z = (1, x_1, x_2, x_1^2, x_2^2, x_1 x_2)$$

$\downarrow$ SVM (Z Space)

Become Linearly Separable

$$(w, b)$$

Objective

$$\min_{\alpha_1,\dots,\alpha_n} \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m \underline{Z_n^T Z_m} - \sum_{n=1}^{N}\alpha_n$$

$$\underline{X_n} \longrightarrow \underline{Z_n}$$

$$\underline{X_m} \longrightarrow \underline{Z_m}$$

$$\sum_{n=1}^{N}\alpha_n y_n = 0 \qquad 0 \leq \alpha_n < \infty$$

$$\underline{W^*} = \sum_{\forall n:\alpha_n>0} \alpha_n y_n \underline{Z_n}$$

$$b = \frac{1}{y_n}(1 - y_n \underline{W^T Z_n})$$

Z Vector Grows Exponentially (Dimensionality)

Issue, After the non linear transformation, dimensionality of $\underline{Z}$ can be very large (takes a long time to train)

Dual Problem

$K(\underline{X_n, X_m})$

Objective $$\sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m \underline{Z_n^T Z_m} - \sum_{n=1}^{N}\alpha_n$$

Constraint $$\sum_{n=1}^{N}\alpha_n y_n = 0 \qquad 0 \leq \alpha_n < \infty$$

$$\underline{X_n} \longrightarrow \underline{Z_n}$$

$$\underline{X_n, X_m} \longrightarrow \underline{Z_n^T Z_m}$$

$K(\underline{x_n, x_m}) = \underline{z_n^T z_m}$

$\downarrow$

Kernel Function

$\alpha_1, \alpha_2, \ldots, \alpha_N$

$$\underline{w} = \sum_{n:\alpha_n>0} \alpha_n y_n \underline{z_n}$$

$$\hat{y} = \text{sign}(\underline{w^T z} + b)$$

$$= \text{sign}(\sum_{n:\alpha_n>0} \alpha_n y_n \underline{z_n^T z} + b)$$

$$= \text{sign}(\sum_{n:\alpha_n>0} \alpha_n y_n K(\underline{x_n, x}) + b)$$

Given a support vector $(\underline{z_m}, y_m)$

$$b = \frac{1}{y_m}(1 - y_m \underline{w^T z_m})$$

$$= \frac{1}{y_m}(1 - y_m \sum_{n:\alpha_n>0} \alpha_n y_n K(\underline{x_n, x_m}))$$

Same Computation Cost As Original Space

We don't need to compute $\underline{z}$ vector

We can go to higher dimensions

How to compute Kernel function $K(\cdot)$?

Example

$X = (x_1, x_2)$

$X^T = (x_1^T, x_2^T)$

$X \longrightarrow Z = (1 \;\; \sqrt{2}x_1 \;\; \sqrt{2}x_2 \;\; \sqrt{2}x_1x_2 \;\; x_1^2 \;\; x_2^2)$

$X' \longrightarrow Z' = (1 \;\; \sqrt{2}x_1' \;\; \sqrt{2}x_2' \;\; \sqrt{2}x_1'x_2' \;\; x_1'^2 \;\; x_2'^2)$

$Z^T Z' = 1 + \underset{2}{x_1x_1'} + \underset{2}{x_2x_2'} + \underset{2}{x_1x_1'x_2x_2'} + x_1^2 x_1'^2 + x_2^2 x_2'^2$

$= (1 + x_1x_1' + x_2x_2')^2$

$= (1 + X^T X)^2$

Qth order transformation

Kernel Function

$K(X, X') = (1 + X^T X')^Q$

Much Faster To Compute

Substitute to dual formulation

Support Vector ↓ ↑ Quality

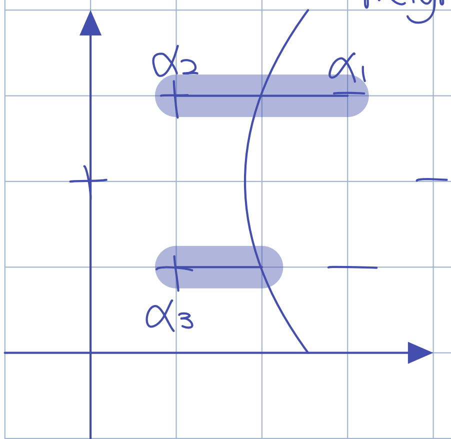
Radial Basis Function ($Q = \infty$)

$$\underline{K(x_n, x_m)} = e^{\overset{\downarrow}{-\gamma} \|\boxed{x_n - x_m}\|^2}$$

Inverse Kernel Width

$$\hat{y} = \text{sign}\Big(\sum_{n:\alpha_n>0} \overset{\downarrow}{\alpha_n} y_n e^{-\gamma \|\boxed{x_n - x}\|^2} + b \Big)$$

Weight Of SV



$d=1,\ \gamma=1$

$$K(x, x') = e^{-(x-x')^2}$$

$$= e^{-(x^2+x'^2-2xx')}$$

$$= e^{-x^2} e^{-x'^2} e^{2xx'}$$

$$e^{\alpha} = 1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^2}{3!} + \ldots$$

$$= e^{-x^2} e^{-x'^2} \left[1 + 2xx' + \frac{(2xx')^2}{2} + \ldots \right]$$

$$= e^{-x^2} e^{-x'^2} \left[1 + (\sqrt{2}x)(\sqrt{2}x') + \frac{(\sqrt{2}x)^2(\sqrt{2}x')^2}{\sqrt{2!}\,\sqrt{2!}} + \frac{(\sqrt{2}x)^3(\sqrt{2}x')^3}{\sqrt{3!}\,\sqrt{3!}} + \ldots \right]$$

$$x \longrightarrow \underline{z} = (z_1, z_2, z_3, \ldots)$$

$$x' \longrightarrow \underline{z'} = (z'_1, z'_2, z'_3, \ldots)$$

$$z^T z' = e^{-x^2} e^{-x'^2} \left[1 + (\sqrt{2}x)(\sqrt{2}x') + \frac{(\sqrt{2}x)^2 (\sqrt{2}x')^2}{\sqrt{2!}\ \sqrt{2!}} + \frac{(\sqrt{2}x)^3 (\sqrt{2}x')^3}{\sqrt{3!}\ \sqrt{3!}} + \dots \right]$$

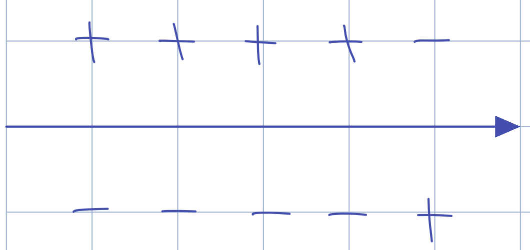
$$z_1 = e^{-x^2}\begin{bmatrix}1\end{bmatrix} \qquad z_2 = e^{-x^2}\begin{bmatrix}\sqrt{2}x_1\end{bmatrix} \qquad \dots$$

$$z_1' = e^{-x'^2}\begin{bmatrix}1\end{bmatrix} \qquad z_2' = e^{-x'^2}\begin{bmatrix}\sqrt{2}x_1'\end{bmatrix}$$

Very Powerful Technic

Soft Margin SVM

How to apply SVM when there are outliers with linearly separable dataset?



Original SVM Formulation

$$\min \frac{1}{2} w^T w \quad \xrightarrow{\quad} \text{Not Feasible}$$

Subject

$$y_n(w^T x_n + b) \geq 1 \quad \xleftarrow{\text{Relax Constraint}}$$

Slack Variable

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

$$\xi_n \geq 0$$

$\xi_n = 0$, Original Problem

Minimize

$$\frac{1}{2}\underline{w}^T\underline{w} + C\sum_{n=1}^{N}\xi_n$$

$$y_n(\underline{w^T x_n} + b) \geq 1 - \xi_n$$

$$\xi_n \geq 0$$

$\xi_n > 1 \longrightarrow (\underline{x_n}, y_n)$ will be misclassified

Can have misclassification and closer margin

$$\xi_n \geq 1 - y_n(\underline{w^T x_n} + b)$$

$$\xi_n \geq 0$$

$$\xi_n \geq \max(0, 1 - y_n(\underline{w^T x_n} + b))$$

$$\underset{\underline{w},\, b}{\text{minimize}} \quad \underbrace{\frac{1}{2C}\underline{w}^T\underline{w}}_{\text{Regularization}} + \underbrace{\sum_{n=1}^{N}\max(0, 1 - y_n(\underline{w^T x_n} + b))}_{e_n(\underline{w})\ \text{“Hinge Loss”}}$$

$$\frac{1}{2C} = \lambda$$

Dual Formulation For Soft Margin SVM

Lagrangian

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

Dual Variables

$$\{\ \alpha_n, \beta_n\ \}_{n=1}^{N}$$

$$\frac{\partial\mathcal{L}}{\partial\xi_n} = C - \alpha_n - \beta_n = 0$$

$$\alpha_n + \beta_n = C \qquad \forall\ n = 1, 2, \ldots, N$$

Lagrangian

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

Dual Problem

$$\underset{\alpha_1,\ldots,\alpha_N}{\text{minimize}}\quad \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m x_n^T x_m - \sum_{n=1}^{N}\alpha_n$$

$$\sum_{n=1}^{N}\alpha_n y_n = 0 \qquad 0 \leq \alpha_i \leq C$$