Support Vector Machine

FMMS050 ML PhD course

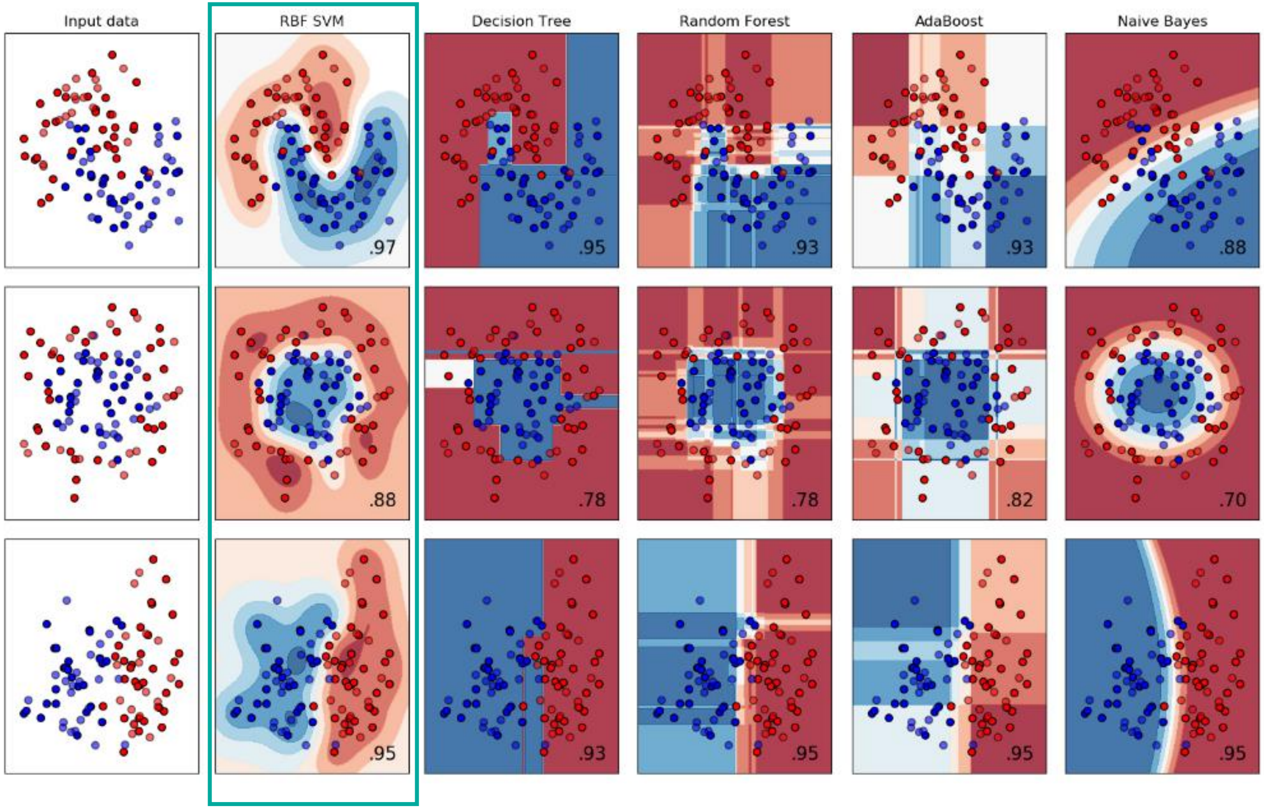
Lecture 8

Support Vector Machine

Most powerful machine learning algorithm before ANN appears

- Support Vector Classification
- Support Vector Regression
- Bayesian SVM

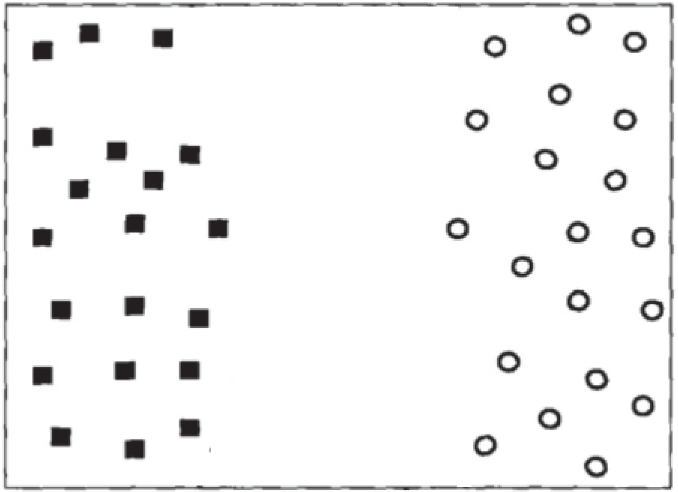
Performance of SVM



Performance of SVM

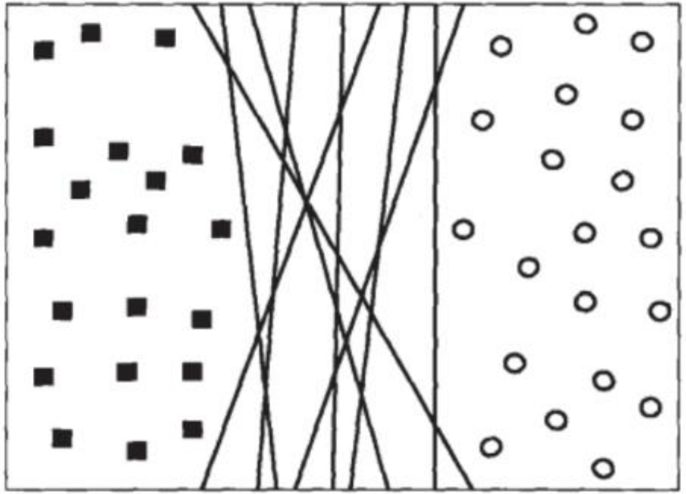
- Linear SVM = a neuron
- Non-linear SVM = two layers of neural network
- SVM is easy, because you may not understand its theory
- SVM is exetremly hard, because you are trying to understand its theory

Hyper plane in SVM



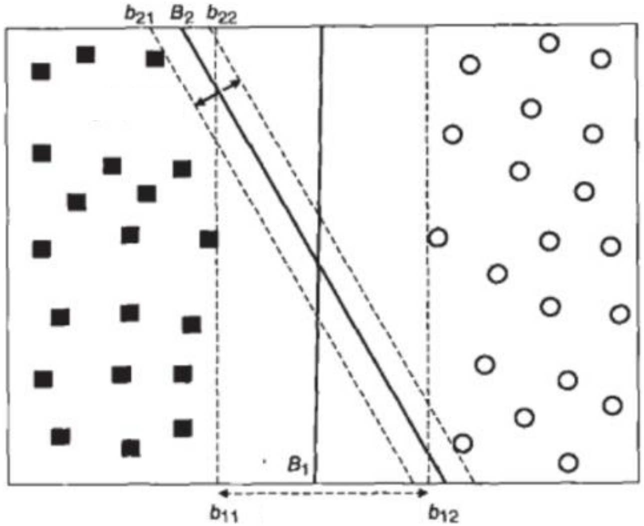
Hyper plane in SVM

- Dim. of hyper palne = dim. of features – 1
- Decision boundary



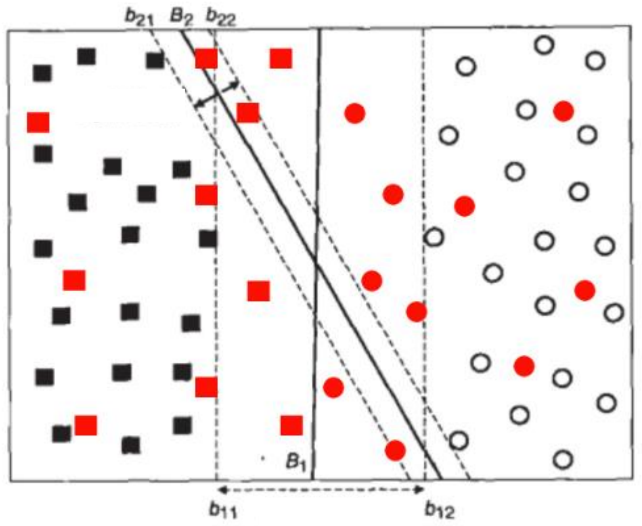
Hyper plane in SVM

- d = b1 – b2: margin
- Virtual decision boundaries



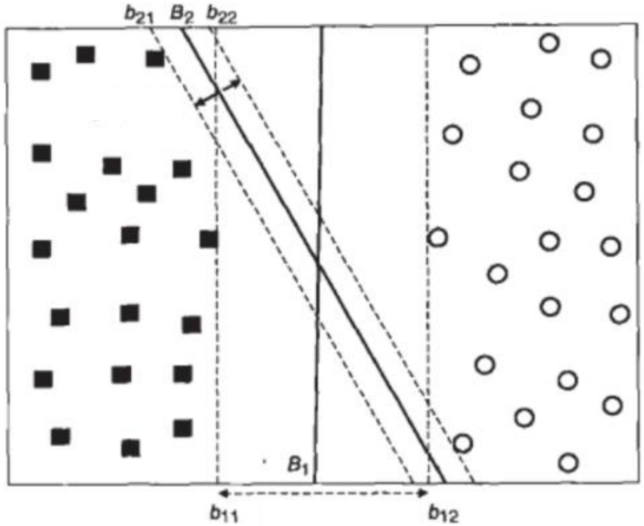
Hyper plane in SVM

- B2, small margin
- B1, large margin
- B2: overfitting
- B1: good-fitting (maxium margin)



What does SVM do?

- Find min of loss function
- Find max of margin



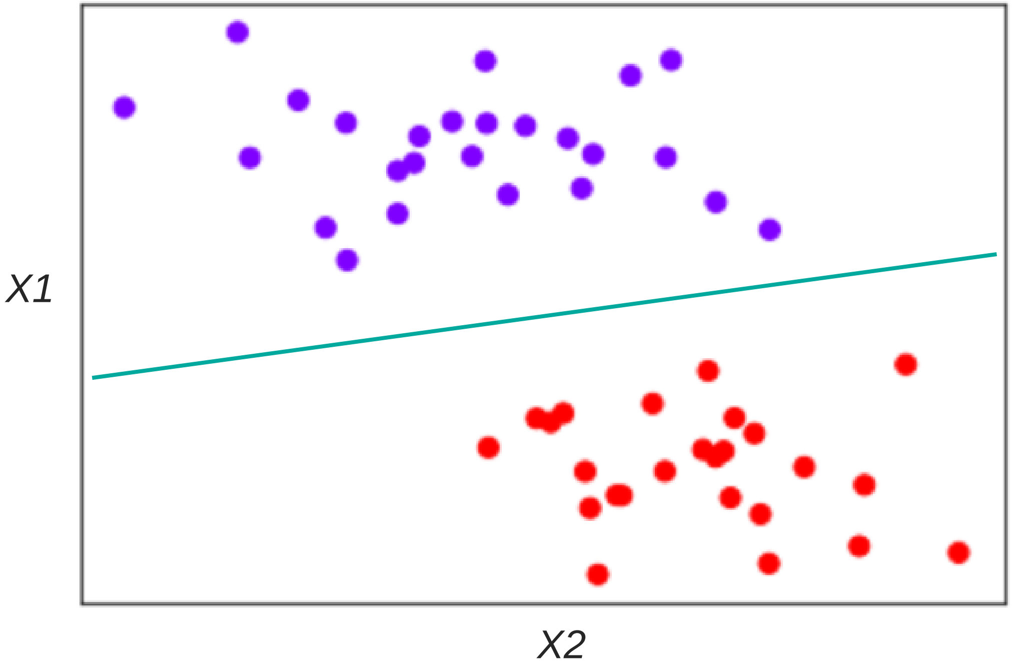
SVM for classification

$$x_1 = ax_2 + b$$

$$0 = ax_2 - x_1 + b$$

$$0 = [a, -1] * \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} + b$$

$$0 = \boldsymbol{w}^T \boldsymbol{x} + b$$



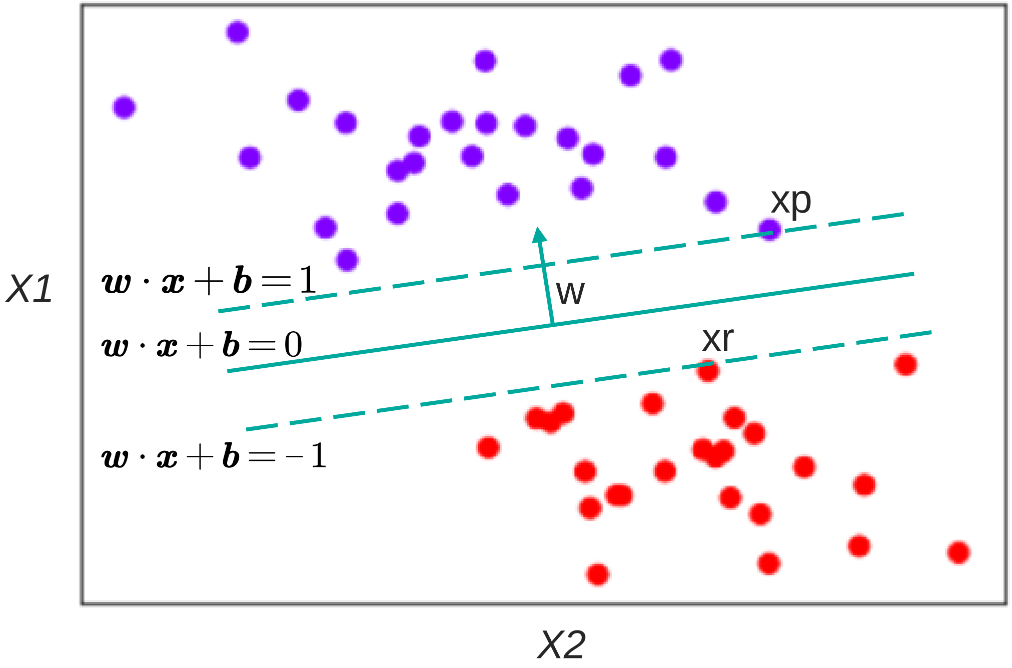
SVM for classification

$$x_1 = ax_2 + b$$

$$0 = ax_2 - x_1 + b$$

$$0 = [a, -1] * \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} + b$$

$$0 = \boldsymbol{w}^T \boldsymbol{x} + b$$



SVM for classification

$$\boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b} = 1$$

$$\boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b} = -1$$

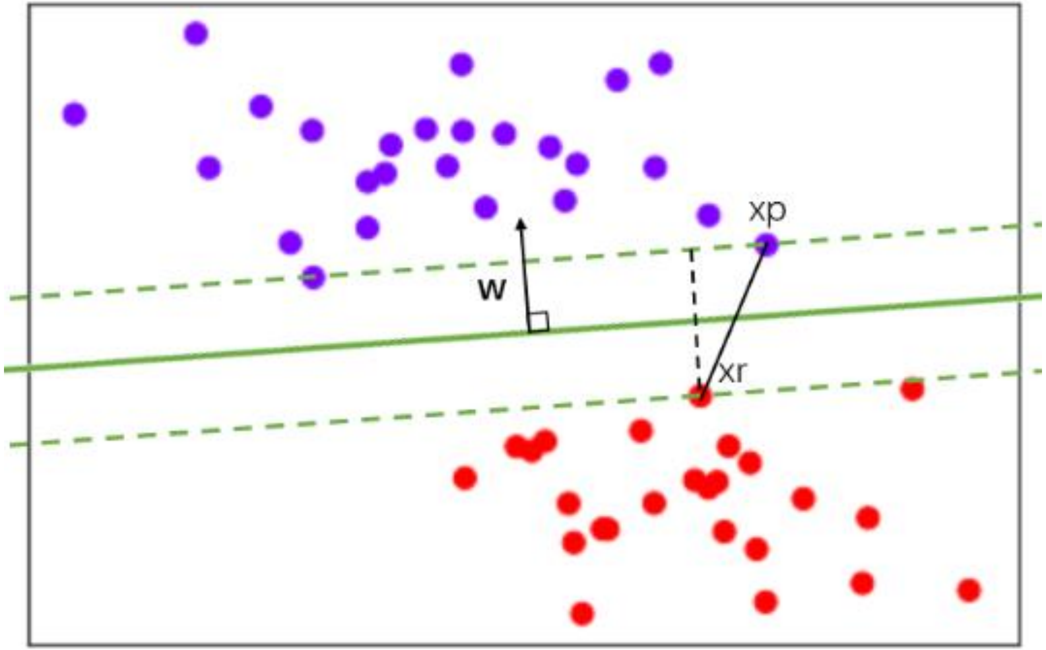
$$\boldsymbol{w} \cdot (\boldsymbol{x_p} - \boldsymbol{x_r}) = 2$$

$$\frac{\boldsymbol{w} \cdot (\boldsymbol{x_p} - \boldsymbol{x_r})}{||w||} = \frac{2}{||w||}$$

$$\therefore \; d = \frac{2}{||w||}$$

max(*d*) = min(*1/d*)

$$f(w) = \frac{||w||^2}{2}$$



SVM for classification

$$if\ \boldsymbol{y} = 1,\ \boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b} \geqslant 1$$

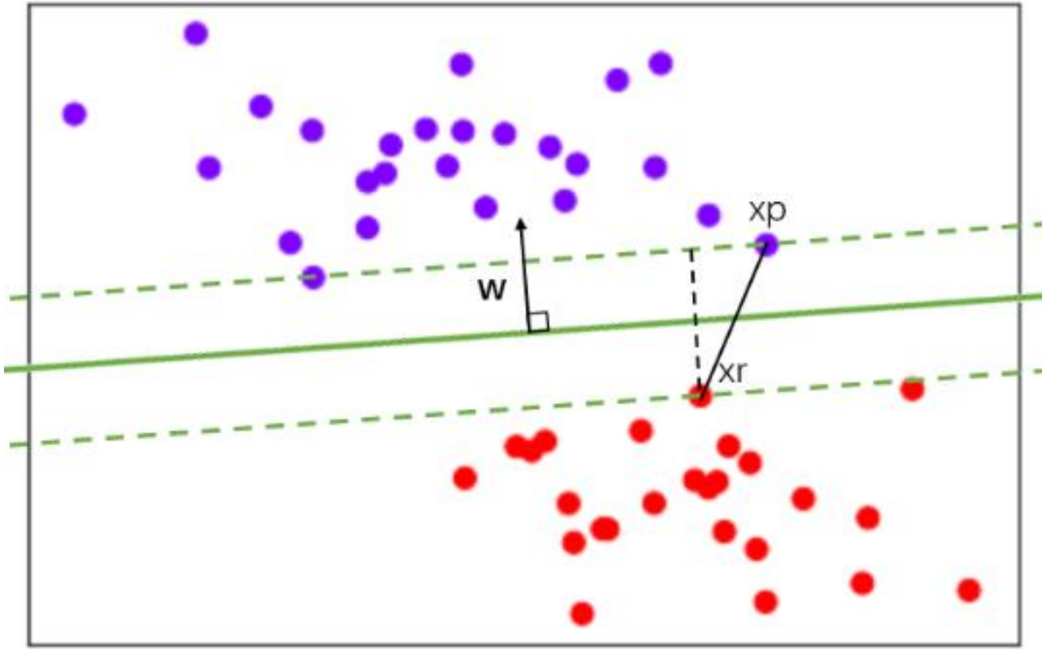
$$if\ \boldsymbol{y} = -1,\ \boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b} \leqslant -1$$

$$\boldsymbol{y} \cdot (\boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b}) \geqslant 1$$

Loss function of linear SVM

$$\min_{w,b} \frac{||\boldsymbol{w}||^2}{2}$$

s.t. $y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) \geq 1,\ \ i = 1, 2, \ldots N$



Constrained minimization

Unconstrained Minimization

$$\boxed{\min \; f(x)}$$

$$\boxed{\nabla_{\mathbf{x}} f(\mathbf{x}^*) = \mathbf{0}}$$

Equality Constraints

$$\boxed{\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) \;\; \text{subject to} \;\; h(\mathbf{x}) = 0}$$

Define the **Lagrangian** as

$$\boxed{\mathcal{L}(\mathbf{x}, \mu) = f(\mathbf{x}) + \mu\, h(\mathbf{x})}$$

Then $\mathbf{x}^*$ a local minimum $\iff$ there exists a unique μ^* s.t.

1. $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \mu^*) = \mathbf{0}$
2. $\nabla_{\mu} \mathcal{L}(\mathbf{x}^*, \mu^*) = 0$

The constrained optimization problem is

$$\boxed{\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) \;\; \text{subject to} \;\; h_i(\mathbf{x}) = 0 \;\; \text{for } i = 1, \ldots, l}$$

Construct the **Lagrangian** (introduce a multiplier for each constraint)

$$\boxed{\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{l} \mu_i\, h_i(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\mu}^t\, \mathbf{h}(\mathbf{x})}$$

Lagrangian problem

Loss function of linear SVM

$$\min_{w,b} \frac{||\boldsymbol{w}||^2}{2}$$

s.t. $y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) \geq 1, \quad i = 1, 2, \ldots N$

The constrained optimization problem is

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \text{ subject to } h_i(\mathbf{x}) = 0 \text{ for } i = 1, \ldots, l$$

Construct the **Lagrangian** (introduce a multiplier for each constraint)

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{l} \mu_i h_i(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\mu}^t \mathbf{h}(\mathbf{x})$$

Lagrangian function

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{N} \alpha_i (y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) - 1) \quad (\alpha_i \geq 0)$$

$$\min_{w,b} \max_{\alpha_i \geq 0} L(w, b, \alpha) \quad (\alpha_i \geq 0)$$

Lagrangian dual problem

$$\min_{w,b} \max_{\alpha_i \geq 0} L(w,b,\alpha) \quad (\alpha_i \geq 0)$$

$$L(w,b,\alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{N} \alpha_i (y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) - 1) \quad (\alpha_i \geq 0)$$

$$\begin{aligned}
L(w,b,\alpha) &= \frac{1}{2}||w||^2 - \sum_{i=1}^{N} \alpha_i (y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) - 1) \\
&= \frac{1}{2}||w||^2 - \sum_{i=1}^{N} (\alpha_i y_i \boldsymbol{w} \cdot \boldsymbol{x_i} + \alpha_i y_i b - \alpha_i) \\
&= \frac{1}{2}||w||^2 - \sum_{i=1}^{N} (\alpha_i y_i \boldsymbol{w} \cdot \boldsymbol{x_i}) - \sum_{i=1}^{N} \alpha_i y_i b + \sum_{i=1}^{N} \alpha_i \\
&= \frac{1}{2}(\boldsymbol{w^T} \boldsymbol{w})^{\frac{1}{2} * 2} - \sum_{i=1}^{N} (\alpha_i y_i \boldsymbol{w} \cdot \boldsymbol{x_i}) - \sum_{i=1}^{N} \alpha_i y_i b + \sum_{i=1}^{N} \alpha_i \\
&= \frac{1}{2}\boldsymbol{w^T} \boldsymbol{w} - \sum_{i=1}^{N} (\alpha_i y_i \boldsymbol{w} \cdot \boldsymbol{x_i}) - \sum_{i=1}^{N} \alpha_i y_i b + \sum_{i=1}^{N} \alpha_i
\end{aligned}$$

$$\begin{aligned}
\frac{\partial L(\boldsymbol{w},b,\alpha)}{\partial \boldsymbol{w}} &= \frac{1}{2} * 2\boldsymbol{w} - \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x_i} \\
&= \boldsymbol{w} - \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x_i} = 0 \rightarrow \boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x_i} \qquad (1)
\end{aligned}$$

$$\frac{\partial L(\boldsymbol{w},b,\alpha)}{\partial b} = \sum_{i=1}^{N} \alpha_i y_i = 0 \rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \qquad (2)$$

Lagrangian dual problem

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{N} \alpha_i (y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) - 1) \ \ (\alpha_i \geq 0)$$

Lagrangian function:

$$L(x, \alpha) = f(x) + \sum_{i=1}^{q} \alpha_i h_i(x)$$

Lagrangian problem:

$$\min_x L(x, \alpha)$$

Lagrangian dual problem:

$$\max_\alpha g(\alpha)$$

dual gap:

$$\Delta = \min_x L(x, \alpha) - \max_\alpha g(\alpha)$$

strong duality property

$$\min_x L(x, \alpha) = \max_\alpha g(\alpha)$$

KKT cond.(Karush-Kuhn-Tucker)

$$\begin{aligned} \frac{\partial L}{\partial x_i} &= 0, \ \forall_i = 1, 2, \ldots, d \\ h_i(x) &\leq 0, \ \forall_i = 1, 2, \ldots, q \\ \alpha_i &\geq 0, \ \forall_i = 1, 2, \ldots, q \\ \alpha_i h_i(x) &= 0, \ \forall_i = 1, 2, \ldots, q \end{aligned}$$

KKT condition

KKT

$$\boxed{\begin{aligned} \frac{\partial L}{\partial x_i} &= 0,\ \forall_i = 1,2,\ldots,d \\ h_i(x) &\le 0,\ \forall_i = 1,2,\ldots,q \\ \alpha_i &\ge 0,\ \forall_i = 1,2,\ldots,q \\ \alpha_i h_i(x) &= 0,\ \forall_i = 1,2,\ldots,q \end{aligned}}$$

$$\sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x_i} = \boldsymbol{w} \quad (1)$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \quad (2)$$

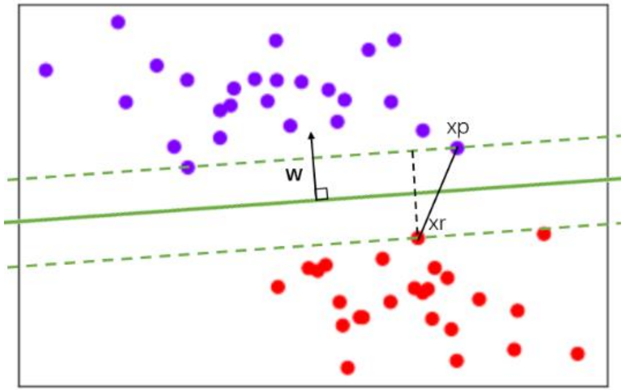
$$-\left(y_i(\boldsymbol{w}\cdot\boldsymbol{x_i}+b)-1\right) \le 0 \quad (3)$$

$$\alpha_i \ge 0 \quad (4)$$

$$\boxed{\alpha_i(y_i(\boldsymbol{w}\cdot\boldsymbol{x_i}+b)-1) = 0 \quad (5)}$$

$$y_i(\boldsymbol{w}\cdot\boldsymbol{x_i}+b)-1=0$$

$$y_i(\boldsymbol{w}\cdot\boldsymbol{x_i}+b)-1=0$$



Get the formulation of dual problem

$$\sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x_i} = \boldsymbol{w} \quad (1)$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \quad (2)$$

$$-\,(y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) - 1) \leq 0 \quad (3)$$

$$\alpha_i \geq 0 \quad (4)$$

$$\alpha_i(y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) - 1) = 0 \quad (5)$$

$$\begin{aligned}
L(w,b,\alpha) &= \frac{1}{2}||w||^2 - \sum_{i=1}^{N} \alpha_i(y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) - 1) \\
&= \frac{1}{2}||w||^2 - \sum_{i=1}^{N}(\alpha_i y_i \boldsymbol{w} \cdot \boldsymbol{x_i}) - \sum_{i=1}^{N} \alpha_i y_i b + \sum_{i=1}^{N} \alpha_i \\
&= \frac{1}{2}||w||^2 - \boldsymbol{w}\sum_{i=1}^{N}(\alpha_i y_i \cdot \boldsymbol{x_i}) - b\sum_{i=1}^{N} \alpha_i y_i + \sum_{i=1}^{N} \alpha_i \\
&= \frac{1}{2}\boldsymbol{w^T w} - \boldsymbol{w^T w} + \sum_{i=1}^{N} \alpha_i \\
&= -\frac{1}{2}\boldsymbol{w^T w} + \sum_{i=1}^{N} \alpha_i \\
&= \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \boldsymbol{x_i^T x_j}
\end{aligned}$$

$$L_d = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \boldsymbol{x_i} \cdot \boldsymbol{x_j}$$

$$\max_{\alpha_i \geq 0}\left(\sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \boldsymbol{x_i} \cdot \boldsymbol{x_j}\right)$$

Solve for parameters

Solve α_i, from $\max_{\alpha_i \geq 0}(\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \boldsymbol{x_i} \cdot \boldsymbol{x_j})$

- GD,
- SMO (sequential minimal optimization),
- quadratic programming.

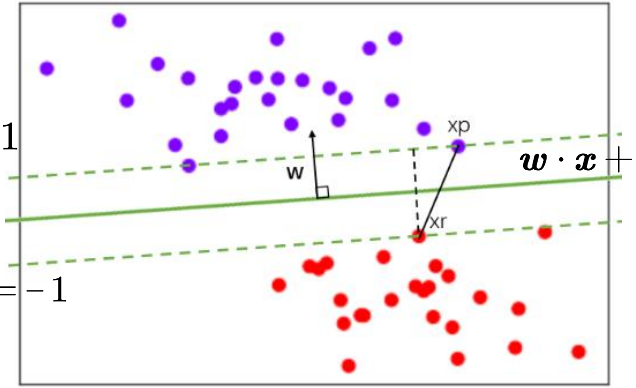
$$\sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x_i} = \boldsymbol{w} \quad (1)$$

$$y_i(w^T x_i + b) = 1$$

$\boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b} = 1$

$\boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b} = 0$

$\boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b} = -1$



Example #1

Kernel function

Mapping from low dim. to high dim.

$$\boxed{x \rightarrow \Phi(x)}$$

Kernel Function

$K(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{\Phi}(\boldsymbol{u}) \cdot \boldsymbol{\Phi}(\boldsymbol{v})$

"linear"	$K(x,y) = x^T y = x \cdot y$
"poly"	$K(x,y) = (\gamma(x \cdot y) + r)^d$
"sigmoid"	$K(x,y) = tanh(\gamma(x \cdot y) + r)$
"rbf"	$K(x,y) = e^{-\gamma\|x-y\|^2}, \gamma > 0$

$$\min_{w,b} \frac{\|\boldsymbol{w}\|^2}{2}$$

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) \geq 1, \quad i = 1, 2, \ldots N$$

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i (y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) - 1) \quad (\alpha_i \geq 0)$$

$$L_d = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \boldsymbol{x_i} \cdot \boldsymbol{x_j}$$

$$\min_{w,b} \frac{\|\boldsymbol{w}\|^2}{2}$$

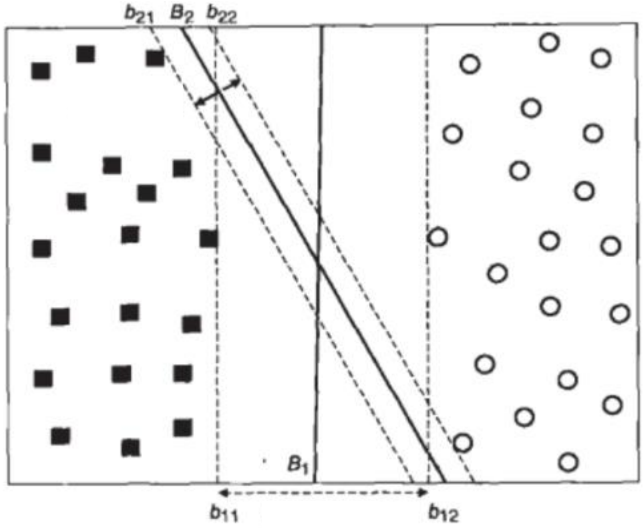
$$\textit{subject to} \quad y_i(\boldsymbol{w} \cdot \boldsymbol{\Phi}(\boldsymbol{x_i}) + b \geq 1), \quad i = 1, 2, \ldots N$$

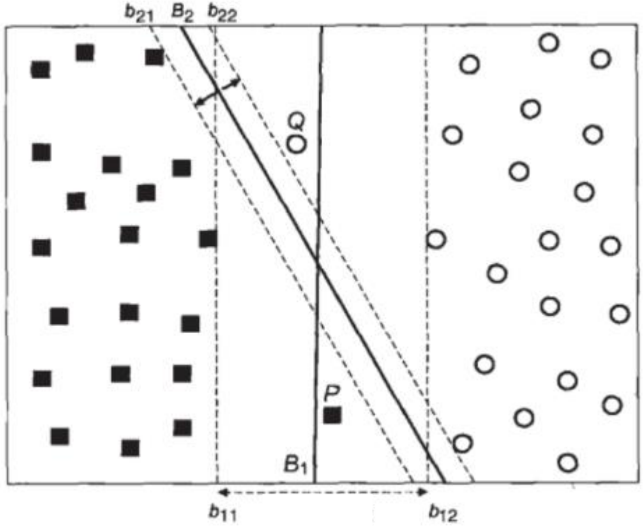
$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i (y_i(\boldsymbol{w} \cdot \boldsymbol{\Phi}(\boldsymbol{x_i}) + b) - 1)$$

$$L_d = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \boxed{\boldsymbol{\Phi}(\boldsymbol{x_i}) \boldsymbol{\Phi}(\boldsymbol{x_j})}$$

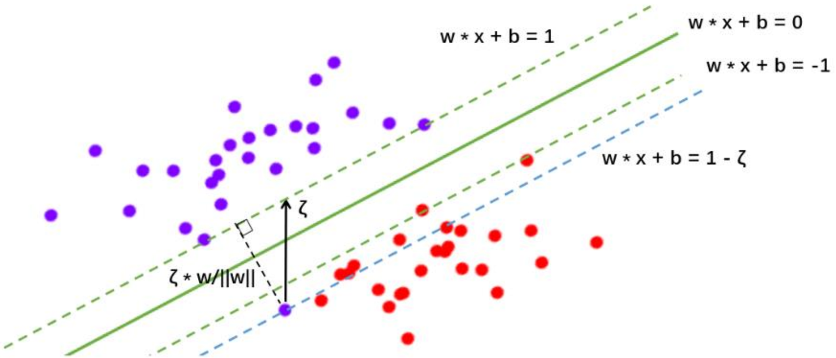
Example #2

Soft margin





Soft margin



New boundaries

$$\boldsymbol{w} \cdot \boldsymbol{x_i} + b \geq 1 - \zeta_i \quad if \;\; y_i = 1$$
$$\boldsymbol{w} \cdot \boldsymbol{x_i} + b \leq -1 + \zeta_i \quad if \;\; y_i = -1$$

New loss function

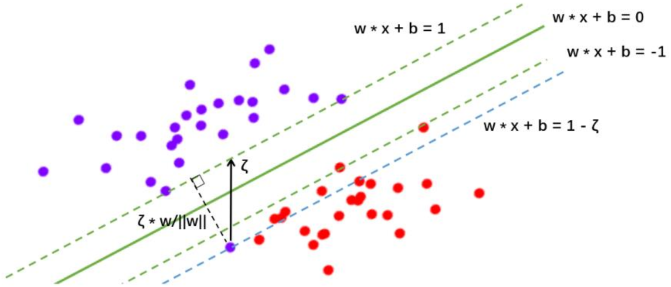
$$\min_{w,b,\zeta} \frac{||\boldsymbol{w}||^2}{2} + C \sum_{i=1}^{n} \zeta_i$$

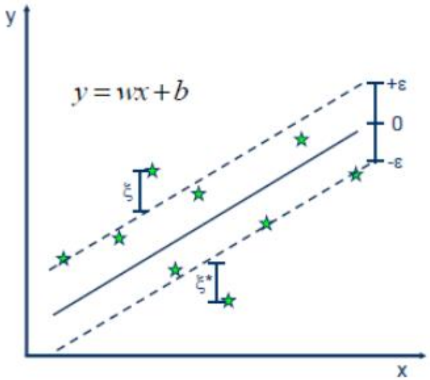
$$\begin{aligned} subject\ to \qquad & y_i(\boldsymbol{w} \cdot \boldsymbol{\Phi}(\boldsymbol{x_i}) + b \geq 1 - \zeta_i), \\ & \zeta_i \geq 0, \\ & i = 1, 2, \ldots N \end{aligned}$$

$$L(w, b, \alpha, \zeta) = \frac{1}{2}||w||^2 + C \sum_{i=1}^{N} \zeta_i - \sum_{i=1}^{N} \alpha_i (y_i(\boldsymbol{w} \cdot \boldsymbol{\Phi}(\boldsymbol{x_i}) + b) - 1 + \zeta_i) - \sum_{i=1}^{N} \mu_i \zeta_i$$

$$\frac{\partial L(w,b,\alpha,\zeta)}{\partial w} = \frac{\partial L(w,b,\alpha,\zeta)}{\partial b} = \frac{\partial L(w,b,\alpha,\zeta)}{\partial \zeta} = 0$$
$$\zeta_i \geq 0, \;\; \alpha_i \geq 0, \;\; \mu_i \geq 0$$
$$\alpha_i(y_i(\boldsymbol{w} \cdot \boldsymbol{\Phi}(\boldsymbol{x_i}) + b) - 1 + \zeta_i) = 0$$
$$\mu_i \zeta_i = 0$$
$$L_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \boldsymbol{\Phi}(\boldsymbol{x_i}) \boldsymbol{\Phi}(\boldsymbol{x_j})$$
$$subject\ to \;\; C \geq \alpha_i \geq 0$$

Support vector regression





- Minimize:

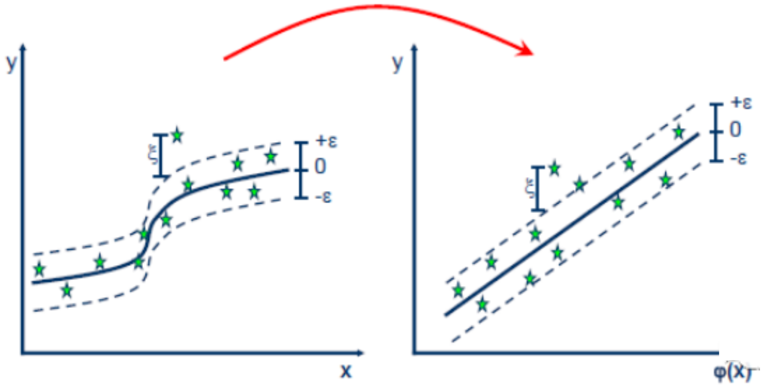
$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right)$$

- Constraints:

$$y_i - wx_i - b \leq \varepsilon + \xi_i$$

$$wx_i + b - y_i \leq \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$



AVANCEZ

1829

CHALMERS

UNIVERSITY OF TECHNOLOGY