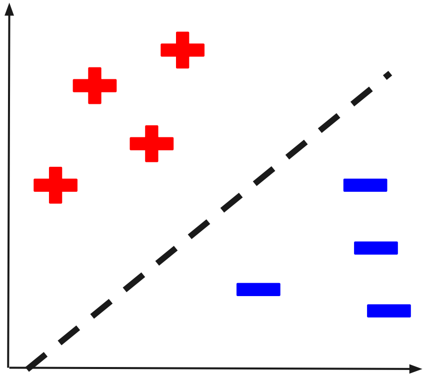
SVM

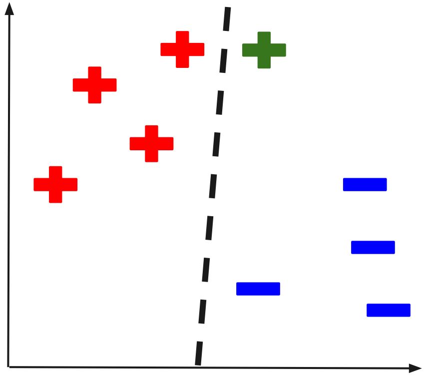
Kang-Wen Wei

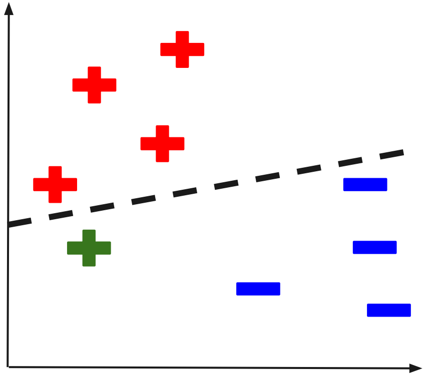
- SVM
  - Linear SVM
    - Decision boundary
    - Vector projection on Margin
    - Lagrange multiplier
  - Kernel function
  - Demo
  - rbf proof

Linear SVM

Which is better ?







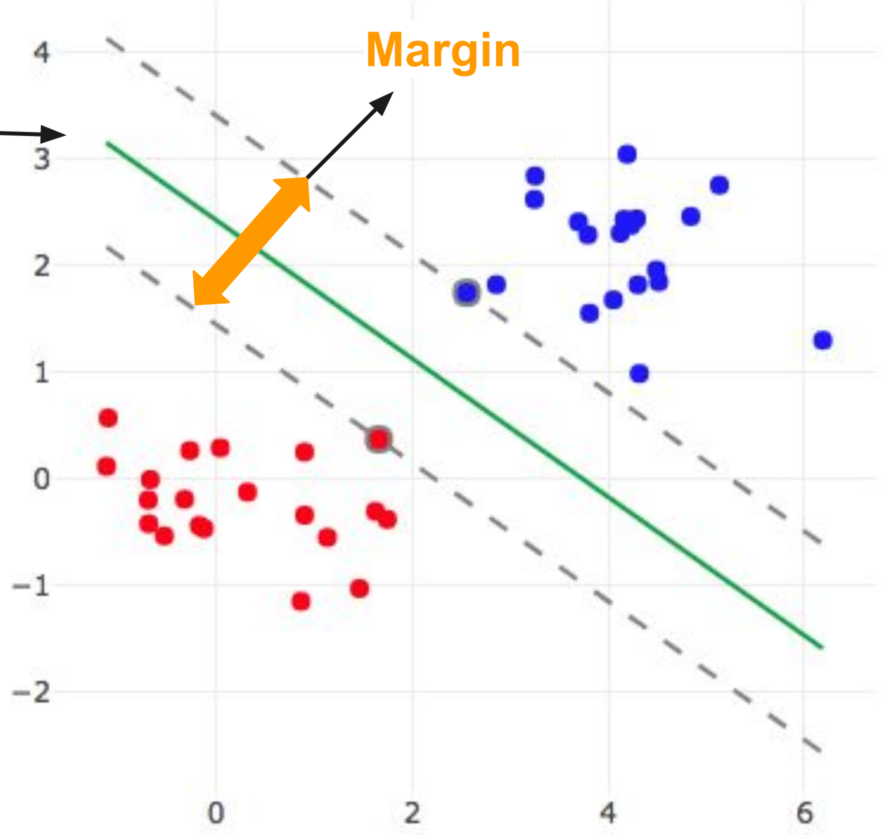
Tolerate more noise !

Decision boundary

Decision boundary

$$y_i = \vec{w}\vec{\mu} + b$$

$\vec{w}$為平面參數，$\vec{\mu}$為$\vec{x}$的向量



1. **Find decision boundary**
2. **Maximum margin**

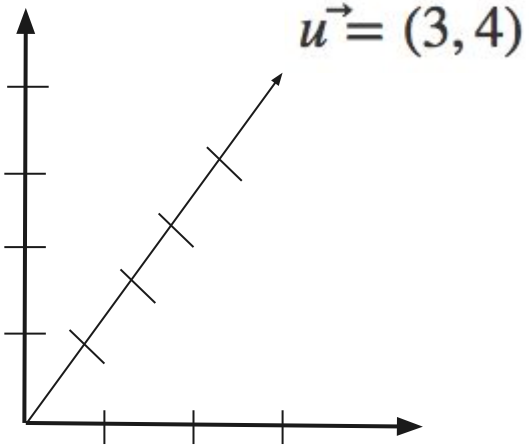
Find Decision boundary

$$y_i = \vec{w}\vec{\mu} + b \Rightarrow \begin{cases} \vec{w} \bullet \vec{x_+} + b \geq 1 \\ \vec{w} \bullet \vec{x_-} + b \leq -1 \end{cases}, \text{such that } y_i = \begin{cases} +1 & \text{, for + examples} \\ -1 & \text{, for -- examples} \end{cases}$$

$$\Rightarrow \begin{cases} y_i(\vec{w}\vec{x_i} + b) \geq 1 \\ y_i(\vec{w}\vec{x_i} + b) \geq 1 \end{cases} \Rightarrow y_i(\vec{w}\vec{x_i} + b) - 1 \geq 0$$

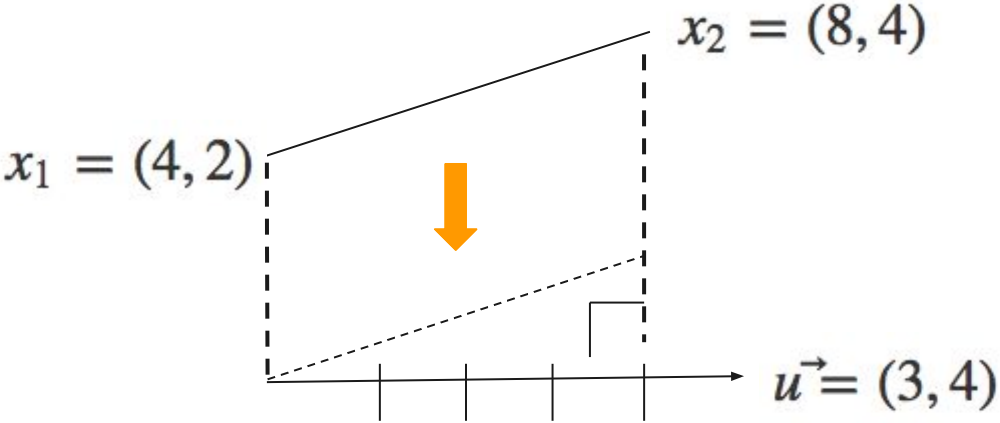
$$\Rightarrow y_i(\vec{w}\vec{x_i} + b) - 1 = 0$$

Unit vector



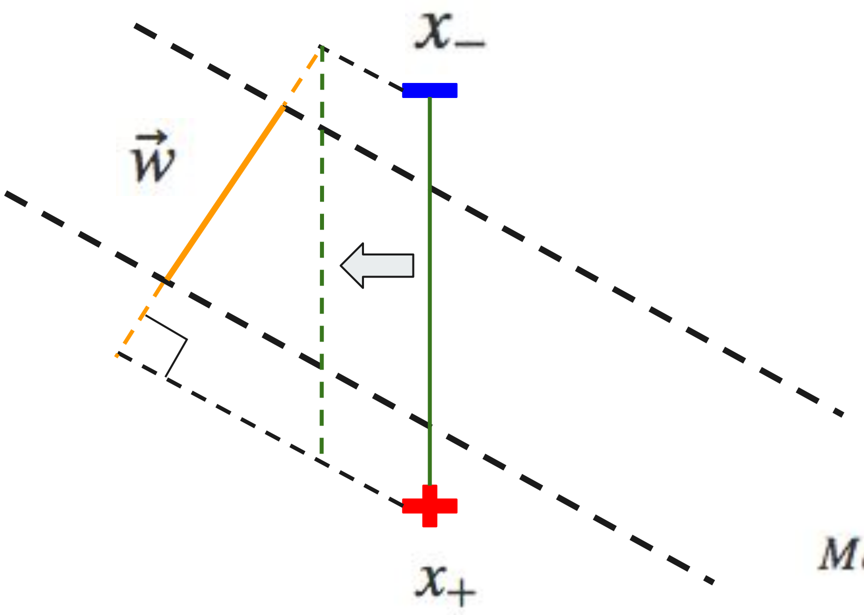
$$\frac{\vec{u}}{||\vec{u}||} = \left(\frac{3}{5}, \frac{4}{5}\right)$$

Vector projection



$$(x_2 - x_1) \bullet \frac{\vec{u}}{||\vec{u}||} = (4,2) \bullet \left(\frac{3}{5}, \frac{4}{5}\right) = 4$$

Maximum margin



$$\text{width} = (\vec{x_+} - \vec{x_-}) \bullet \frac{\vec{w}}{||\vec{w}||}$$

$$= \frac{\vec{x_+}\vec{w} - \vec{x_-}\vec{w}}{||\vec{w}||}$$

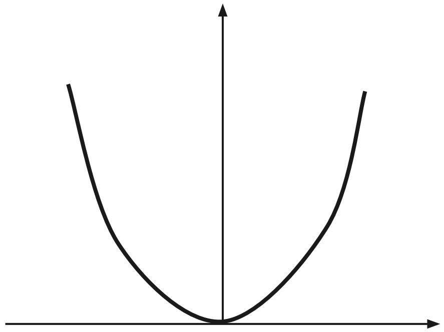
$$= \frac{(1-b)-(-1-b)}{||\vec{w}||}$$

$$= \frac{2}{||\vec{w}||}$$

$$Max\frac{2}{||\vec{w}||} \Rightarrow Max\frac{1}{||\vec{w}||} \Rightarrow Min||\vec{w}|| \Rightarrow Min\frac{1}{2}||\vec{w}||^2$$

Lagrange multiplier

$$Min\ \frac{1}{2}||\vec{w}||^2 \Rightarrow Condition:\ y_i(\vec{w}\overrightarrow{x_i}+b)-1=0$$



$$y = x^2 \Rightarrow \frac{dy}{dx} = 2x = 0$$

$$x = 0 \Rightarrow y = 0$$

$$O : f(x, y) \qquad C : g(x, y)$$

$$L(x, y, \lambda) = f(x, y) - \lambda g(x, y)$$

$$\begin{cases} \dfrac{\partial f}{\partial x} = \lambda \dfrac{\partial g}{\partial x} \\ \dfrac{\partial f}{\partial y} = \lambda \dfrac{\partial g}{\partial y} \\ g = 0 \end{cases} \Rightarrow x, y, \lambda$$

Lagrange multiplier

$$L = \frac{1}{2}||\vec{w}||^2 - \sum \lambda_i[y_i(\vec{w}\vec{x_i} + b) - 1]$$

$$\begin{cases} \frac{\partial L}{\partial \vec{w}} = \vec{w} - \sum \lambda_i y_i \vec{x_i} = 0 \\ \frac{\partial L}{\partial \vec{b}} = \sum \lambda_i y_i = 0,\ \lambda_i \geq 0 \\ \sum \lambda_i[y_i(\vec{w}\vec{x_i} + b) - 1] = 0 \end{cases}$$

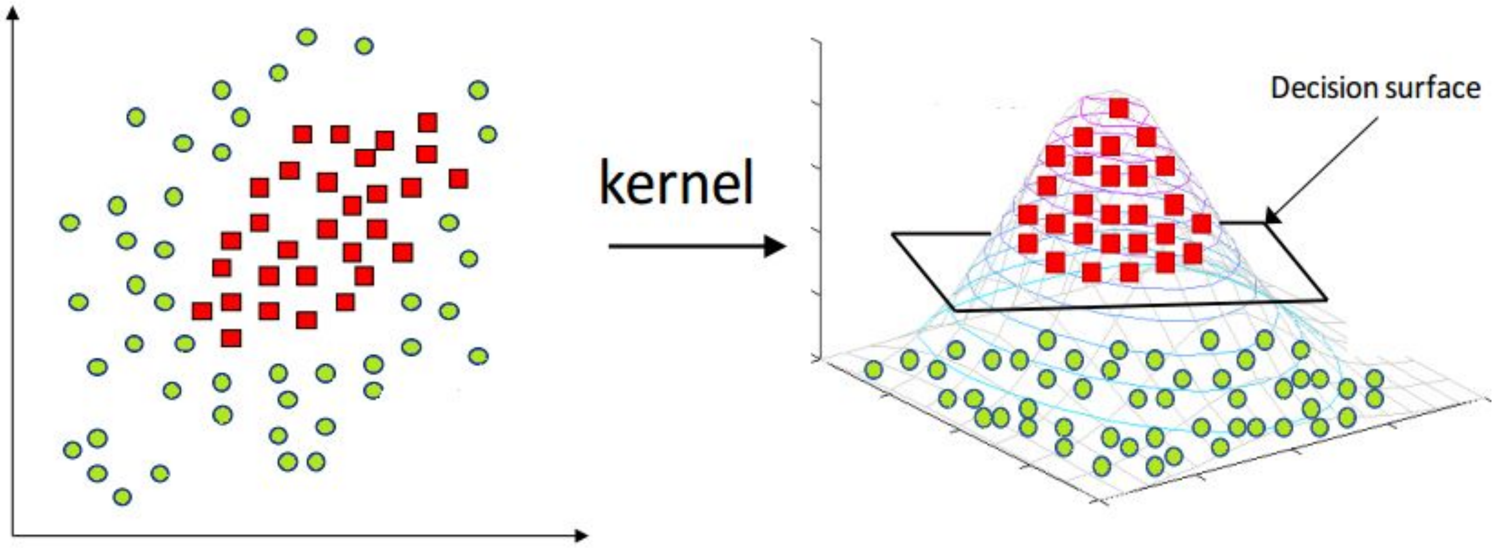
$$\Rightarrow \begin{cases} \vec{w} = \sum \lambda_i y_i \vec{x_i} \\ \sum \lambda_i y_i = 0 \end{cases}$$

$$\begin{aligned} L &= \frac{1}{2}||\vec{w}||^2 - \sum \lambda_i y_i \vec{w}\vec{x_i} - \sum \lambda_i y_i b + \sum \lambda_i \\ &= \sum \lambda_i - \frac{1}{2}||\vec{w}||^2 \\ &= \boxed{\sum \lambda_i - \frac{1}{2}\sum_i \sum_j \lambda_i \lambda_j y_i y_j \vec{x_i}\vec{x_j}\ , \lambda_i \geq 0} \end{aligned}$$

Kernel function

Minimize $\sum \lambda_i - \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j \boxed{\vec{x_i}\,\vec{x_j}}\,, \lambda_i \geq 0$

$$\Rightarrow \sum \lambda_i - \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j \boxed{K(\vec{x_i}\,, \vec{x_j})}\,, \lambda_i \geq 0$$



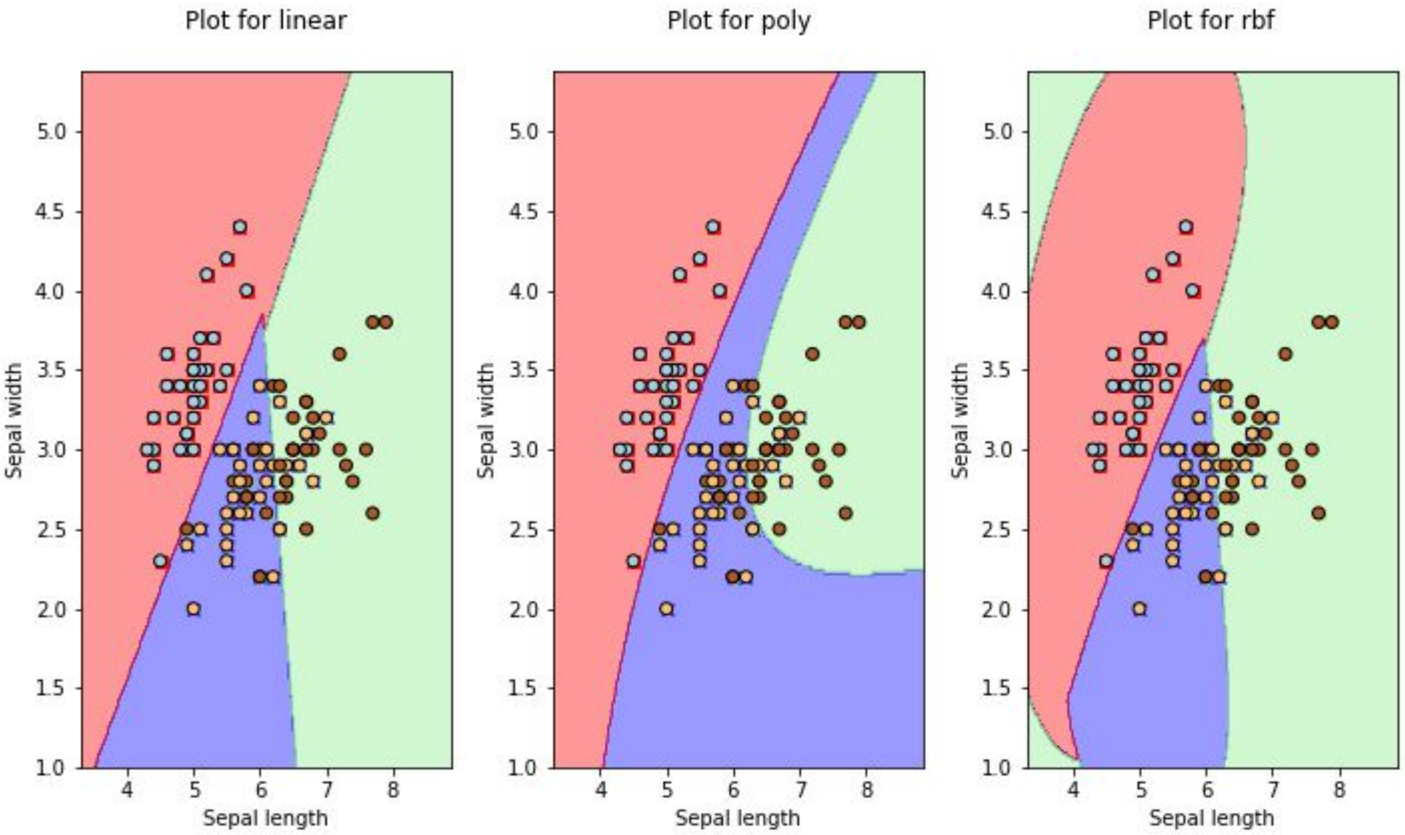
Kernel function

$$Linear : K(x_i, x_j) = x_i^T x_j$$

$$Polynomial : K(x_i, x_j) = (\gamma x_i^T x_j + c)^d, d > 1$$

$$Radial\ basis\ function : K(x_i, x_j) = exp(-\gamma||x_i - x_j||^2), \gamma > 0$$

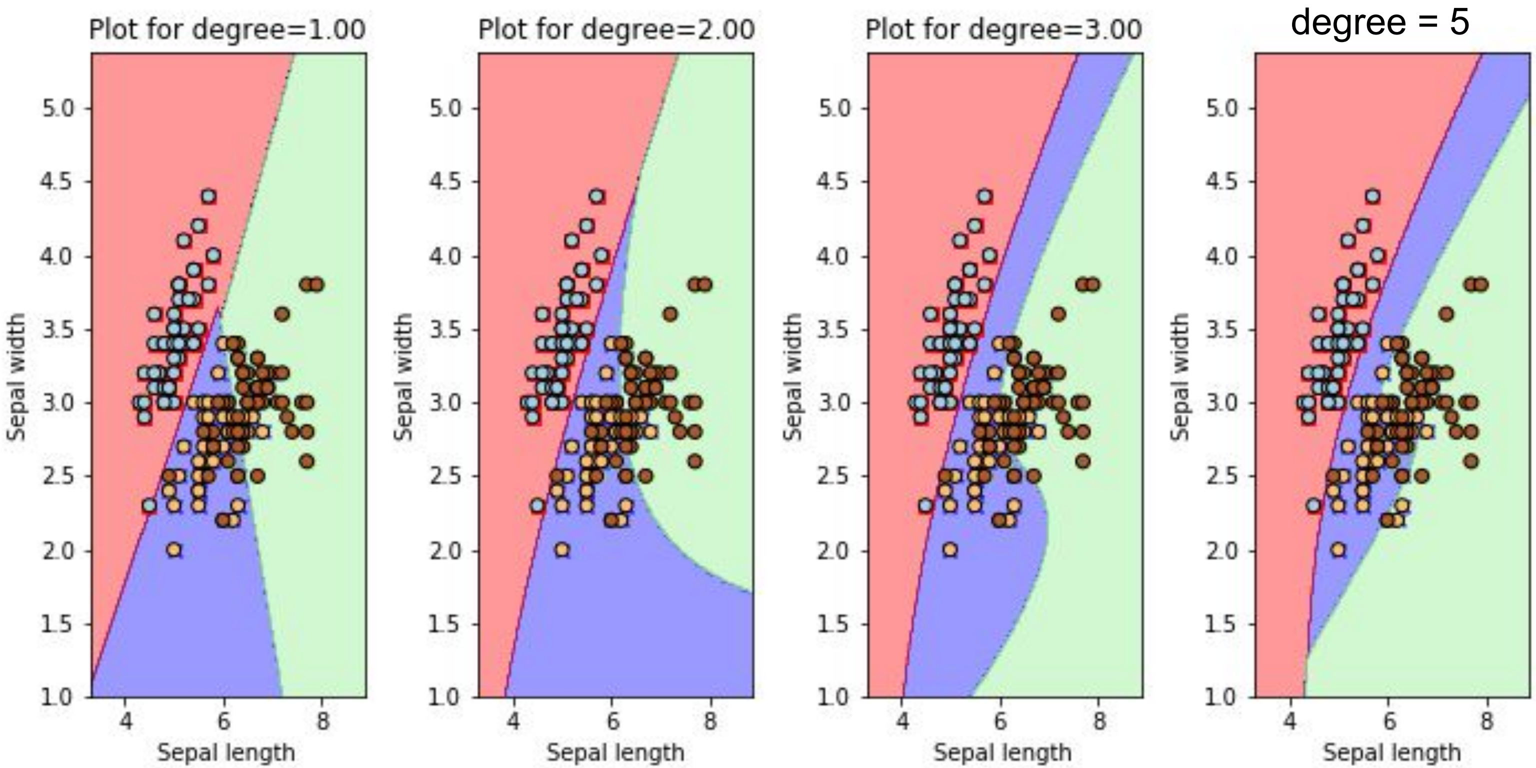
Demo



Plot for linear
Plot for poly
Plot for rbf
Sepal width
Sepal length
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
4
5
6
7
8

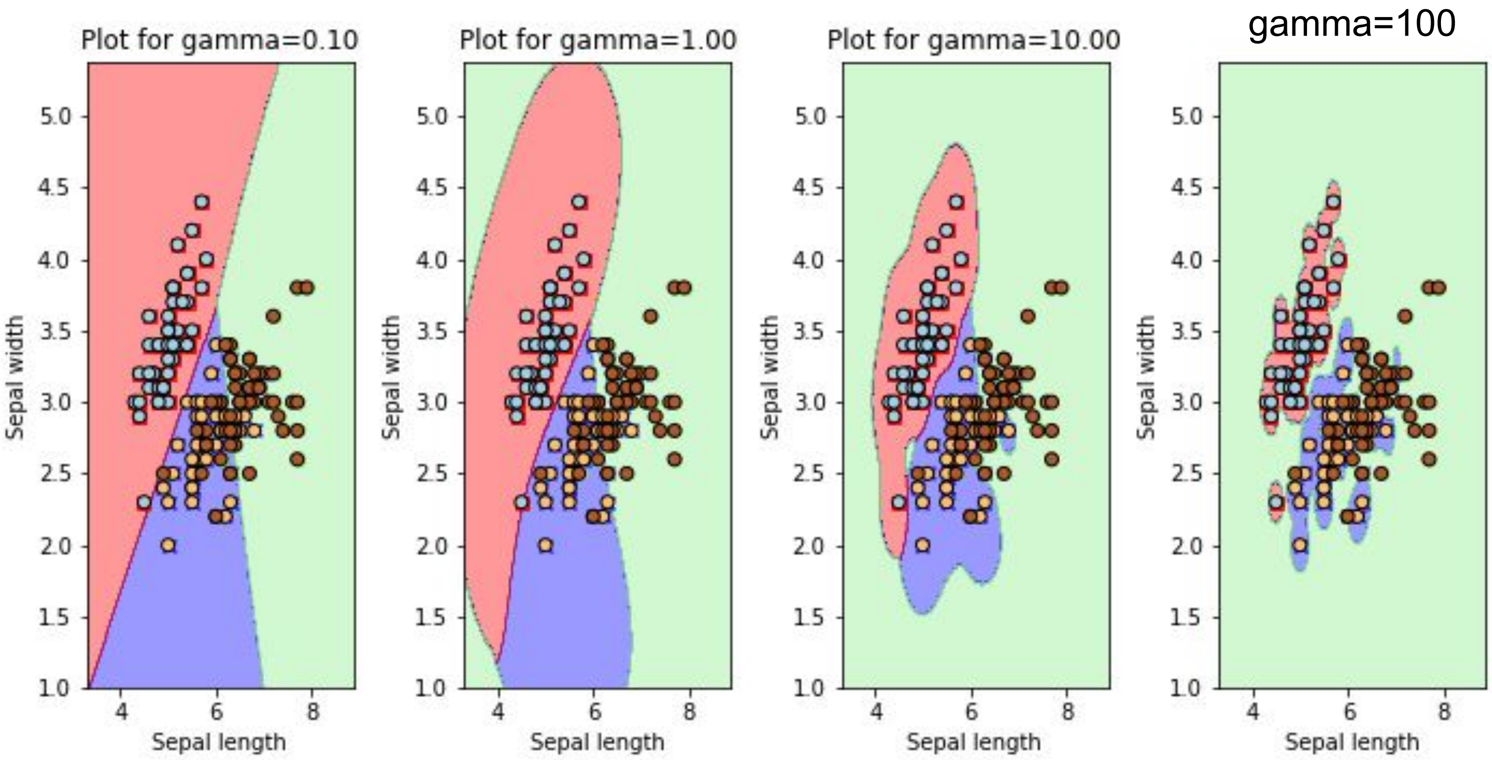
Demo

$$Polynomial : K(x_i, x_j) = (\gamma x_i^T x_j + c)^d, d > 1$$



Demo

$$Radial\ biasis\ function : K(x_i, x_j) = exp(-\gamma||x_i - x_j||^2), \gamma > 0$$



RBF proof

$$Radial\ basis\ function: K(x_i, x_j) = exp(-\gamma||x_i - x_j||^2), \gamma > 0$$

$$K(x_i, x_j) = exp(-\frac{||x_i - x_j||^2}{2\sigma^2}) \longrightarrow \phi^T(x_i) \bullet \phi(x_j)$$

$$\because ||x_i - x_j||^2 = x_i^T x_i + x_j^T x_j - 2x_i^T x_j$$

$$\therefore K(x_i, x_j) = exp(-\frac{x_i^T x_i}{2\sigma^2}) \bullet exp(-\frac{x_j^T x_j}{2\sigma^2}) \bullet \boxed{exp(\frac{x_i^T x_j}{\sigma^2})}$$

Tayler Series

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

$$\begin{aligned} exp\left(\frac{x_i^T x_j}{\sigma^2}\right) &= \sum_{n=0}^{\infty} \frac{1}{n!} \bullet \frac{(x_i^T x_j)^n}{\sigma^{2n}} \\ &= \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} \bullet \frac{(x_i^T)^n}{\sigma^n} \bullet \frac{1}{\sqrt{n!}} \bullet \frac{(x_j^T)^n}{\sigma^n} \\ &= \varphi^T(x_i) \bullet \varphi(x_j) \end{aligned}$$

RBF proof

$$
\begin{aligned}
K(x_i, x_j) &= exp(-\frac{x_i^T x_i}{2\sigma^2}) \bullet exp(-\frac{x_j^T x_j}{2\sigma^2}) \bullet exp(\frac{x_i^T x_j}{\sigma^2}) \\
&= exp(-\frac{x_i^T x_i}{2\sigma^2}) \bullet exp(\frac{x_i^T x_j}{\sigma^2}) \bullet exp(-\frac{x_j^T x_j}{2\sigma^2}) \\
&= exp(-\frac{x_i^T x_i}{2\sigma^2}) \bullet \varphi^T(x_i) \bullet \varphi(x_j) \bullet exp(-\frac{x_j^T x_j}{2\sigma^2}) \\
&= \boxed{\phi^T(x_i) \bullet \phi(x_j)}
\end{aligned}
$$

其中：$\phi^T(x_i) = exp(-\frac{x_i^T x_i}{2\sigma^2}) \bullet \varphi^T(x_i)$

$$\varphi^T(x) = \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} \bullet \frac{(x^T)^n}{\sigma^n}$$