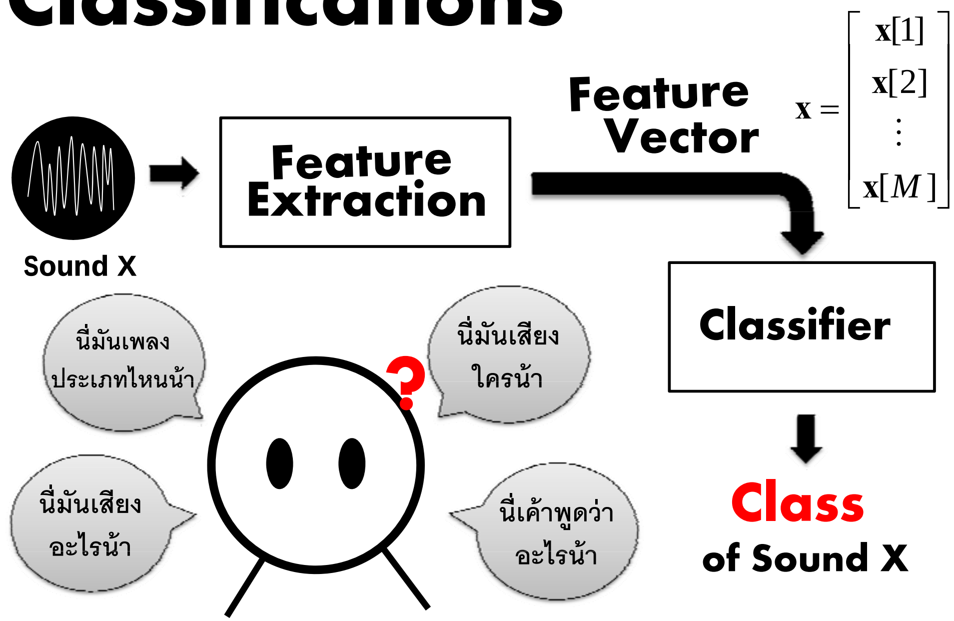
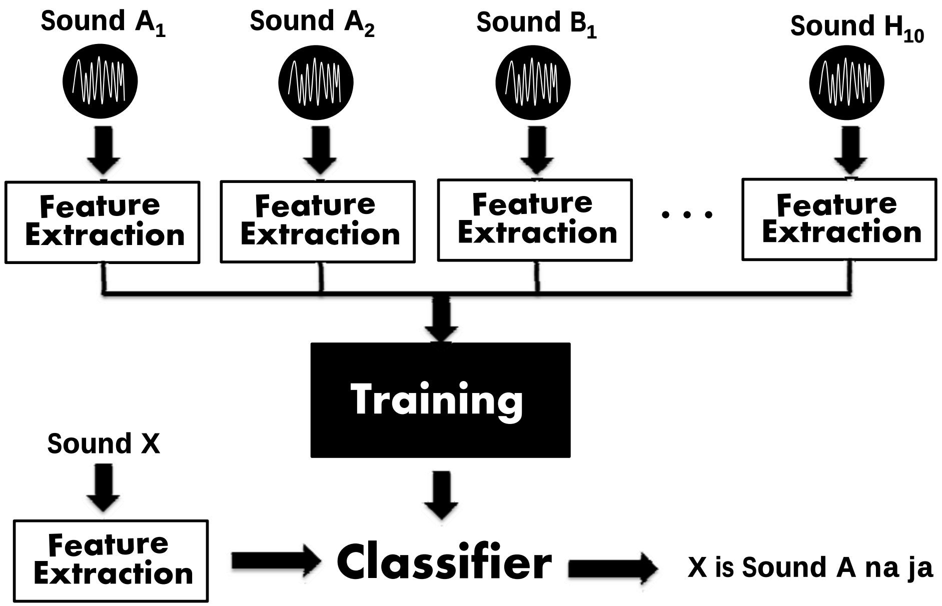
Audio & Speech Technology

[3] Sound Classification

Sound Classifications



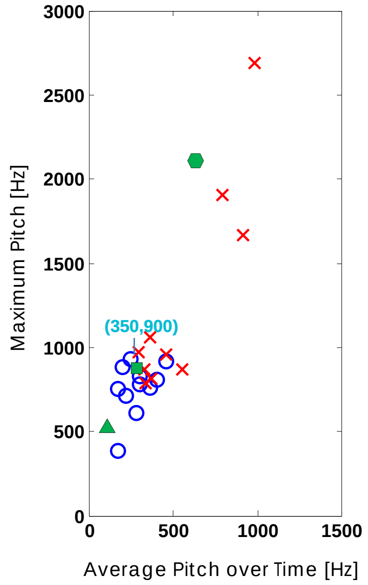
Classifier



Classifier

- k-Nearest Neighbor (kNN)
- Naive Bayes Classifier (NBC)
- Gaussian Mixture Model (GMM)
- Neural Network (NN)
- Support Vector Machine (SVM)
- Hidden Markov Model (HMM)

k-Nearest Neighbor Algorithm



Avg Pitch	Max Pitch	Gender	Distance	Rank
174	385	M	544.2	18
168	755	M	232.7	16
302	780	M	129.2	10
297	830	M	87.8	3
401	805	M	107.8	6
364	760	M	140.7	11
242	930	M	112.1	8
218	710	M	231.4	15
281	610	M	298.1	17
199	880	M	152.3	12
456	915	M	107.3	5
368	815	F	86.9	2
984	2690	F	1899.0	21
791	1905	F	1097.5	20
291	970	F	91.5	4
917	1665	F	952.2	19
552	870	F	204.2	14
333	790	F	111.3	7
363	1060	F	160.5	13
454	955	F	117.6	9
324	870	F	39.7	1

k-Nearest Neighbor Algorithm

Training Set $(N\ Samples)$

$$\{(\mathbf{x}_1, c_1), (\mathbf{x}_2, c_2), (\mathbf{x}_3, c_3) \cdots (\mathbf{x}_N, c_N)\}$$

$\mathbf{x}_i$: *Feature vector of the* i^{th} *sample*

c_i : *Known class of the* i^{th} *sample*

Distance

- **Euclidian Distance**

$$d(\mathbf{a},\mathbf{b}) = \sqrt{\sum_{i=1}^{M} (\mathbf{a}[i] - \mathbf{b}[i])^2}$$

- **Manhattan Distance**

$$d(\mathbf{a},\mathbf{b}) = \sum_{i=1}^{M} |\mathbf{a}[i] - \mathbf{b}[i]|$$

- **Minkowski Distance**
- **Hamming Distance**
- **Levenshtein Distance**

k-Nearest Neighbor Algorithm

Classify Majority vote from k nearest points

$\mathbf{y}$: *Feature vector of unknown class*

⬇

$d(\mathbf{y}, \mathbf{x}_i)$: *Distance between* $\mathbf{y}$ *and* $\mathbf{x}_i$ *for* $i = 1,2,3,.., N$

⬇

$(\mathbf{x}'_i, c'_i)$: *Sample with the* i^{th} *nearest distance* d *for* $i = 1,2,3,..., k$

⬇

$$Class(\mathbf{y}) = Mode(c'_1, c'_2, c'_3, ..., c'_k)$$

Naive Bayes Classifier

Training Set

$$\{(\mathbf{x}_1, c_1), (\mathbf{x}_2, c_2), (\mathbf{x}_3, c_3) \cdots (\mathbf{x}_N, c_N)\}$$

⬇

Probabilistic Model

Feature Vector of Unknown Class

$\mathbf{y}$

⬇

Probabilistic Model

⬇

$Class(\mathbf{y})$

Naive Bayes Classifier

$$Class(\mathbf{y}) = \arg\max_{c}\, p(Class = c \mid F_1 = \mathbf{y}[1], F_2 = \mathbf{y}[2], \ldots, F_M = \mathbf{y}[M])$$

Example

$$P_1 = p(Class = Male \mid Weight = 65kg, Height = 170cm)$$

$$P_2 = p(Class = Female \mid Weight = 65kg, Height = 170cm)$$

$$P_1 > P_2 \Rightarrow Class(65kg, 170cm) = Male$$

$$P_1 < P_2 \Rightarrow Class(65kg, 170cm) = Female$$

Naive Bayes Classifier

Bayes' theorem

$$p(C \mid F_1, F_2, \ldots, F_M) = \frac{p(C)\, p(F_1, F_2, \ldots, F_M \mid C)}{p(F_1, F_2, \ldots, F_M)}$$

Example

$$p(Male \mid W = 65, H = 170) = \frac{p(Male)\, p(W = 65, H = 170 \mid Male)}{p(W = 65, H = 170)}$$

$$p(Male \mid W = 65, H = 170) = \frac{No.of\ males\ with\ 65kg, 170cm}{No.of\ people\ with\ 65kg, 170cm}$$

$$p(W = 65, H = 170 \mid Male) = \frac{No.of\ males\ with\ 65kg, 170cm}{Total\ no.of\ males}$$

$$p(Male) = \frac{Total\ no.of\ males}{Total\ no.of\ people}$$

$$p(W = 65, H = 170) = \frac{No.of\ people\ with\ 65kg, 170cm}{Total\ no.of\ people}$$

Naive Bayes Classifier

$$p(C \mid F_1, F_2, \ldots, F_M) = \frac{p(C)\, p(F_1, F_2, \ldots, F_M \mid C)}{p(F_1, F_2, \ldots, F_M)}$$

$$p(C \mid F_1, F_2, \ldots, F_M) \propto p(C)\, p(F_1, F_2, \ldots, F_M \mid C)$$

Naive conditional independence

F_i, F_j are independence for $i \neq j$

$$p(C \mid F_1, F_2, \ldots, F_M) \propto p(C)\, p(F_1 \mid C)(F_2 \mid C) \cdots (F_M \mid C)$$

Naive Bayes Classifier

Example

$$p(Male \mid W = 65, H = 170) = \frac{p(Male)\, p(W = 65, H = 170 \mid Male)}{p(W = 65, H = 170)}$$

$$p(Male \mid W = 65, H = 170) \propto p(Male)\, p(W = 65 \mid Male)\, p(H = 170 \mid Male)$$

$$p(W = 65, H = 170 \mid Male) = \frac{No.of\ males\ with\ 65kg, 170cm}{Total\ no.of\ males} \Rightarrow \begin{cases} p(W = 65 \mid Male) = \dfrac{No.of\ males\ with\ 65kg}{Total\ no.of\ males} \\[2ex] p(H = 170 \mid Male) = \dfrac{No.of\ males\ with\ 170cm}{Total\ no.of\ males} \end{cases}$$

Naive Bayes Classifier

Features ⇒ Gaussian Distribution

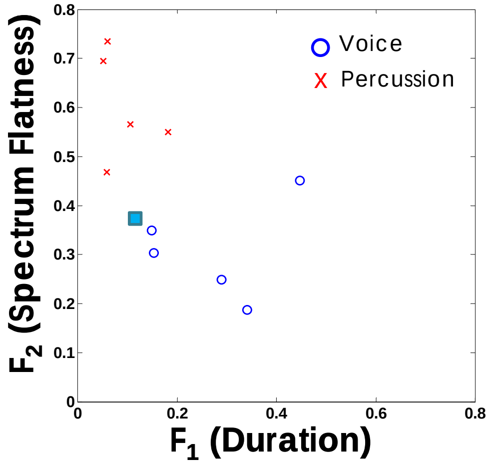
$$f(F=x|Class=c)=\frac{1}{\sqrt{2\pi\sigma^2_{F,c}}}e^{-\frac{(x-\mu_{F,c})^2}{2\sigma^2_{F,c}}}$$

$$\mu_{F,c} = Mean\ of\ Feature\ F\ in\ Class\ c$$

$$\sigma^2_{F,c} = Varience\ of\ Feature\ F\ in\ Class\ c$$

Naive Bayes Classifier

Class	F_1	F_2
V	0.34	0.19
V	0.15	0.30
V	0.45	0.45
V	0.28	0.25
V	0.13	0.35
P	0.11	0.57
P	0.06	0.74
P	0.06	0.47
P	0.18	0.55
P	0.05	0.69
?	0.14	0.37



Naive Bayes Classifier

$$p(Class=V|F_1=0.14,F_2=0.37)$$
$$p(Class=P|F_1=0.14,F_2=0.37)$$

Which one is greater?

$$p(V|F_1=0.14,F_2=0.37)\propto p(V)\,p(F_1=0.14|V)\,p(F_2=0.37|V)$$

$$p(P|F_1=0.14,F_2=0.37)\propto p(P)\,p(F_1=0.14|P)\,p(F_2=0.37|P)$$

Assume that $p(V)=p(P)$

$$p(F_1=0.14|V)\,p(F_2=0.37|V)$$
$$p(F_1=0.14|P)\,p(F_2=0.37|P)$$

Choose the maximum one

Naive Bayes Classifier

Class	F_1	F_2
V	0.34	0.19
V	0.15	0.30
V	0.45	0.45
V	0.28	0.25
V	0.13	0.35
?	0.14	0.37

$$\mu_{F1,V}=0.2700 \qquad \mu_{F2,V}=0.3080$$

$$\sigma^2_{F1,V}=0.0179 \qquad \sigma^2_{F2,V}=0.0098$$

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$p(F_1=0.14|V)\propto\frac{1}{\sqrt{2\pi\times0.0179}}e^{-\frac{(0.14-0.27)^2}{2\times0.0179}}$$
$$\propto 1.8598$$

$$p(F_2=0.37|V)\propto\frac{1}{\sqrt{2\pi\times0.0098}}e^{-\frac{(0.37-0.3080)^2}{2\times0.0098}}$$
$$\propto 3.3122$$

$$p(F_1=0.14|V)\,p(F_2=0.37|V)\propto 1.8598\times3.3122=6.1600$$

Naive Bayes Classifier

Class	F_1	F_2
P	0.11	0.57
P	0.06	0.74
P	0.06	0.47
P	0.18	0.55
P	0.05	0.69
?	0.14	0.37

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mu_{F1,P} = 0.0920 \qquad \mu_{F2,P} = 0.6040$$

$$\sigma^2_{F1,P} = 0.0030 \qquad \sigma^2_{F2,P} = 0.0120$$

$$p(F_1 = 0.14|P) \propto \frac{1}{\sqrt{2\pi \times 0.0030}} e^{-\frac{(0.14-0.0920)^2}{2\times 0.0030}} \propto 4.9611$$

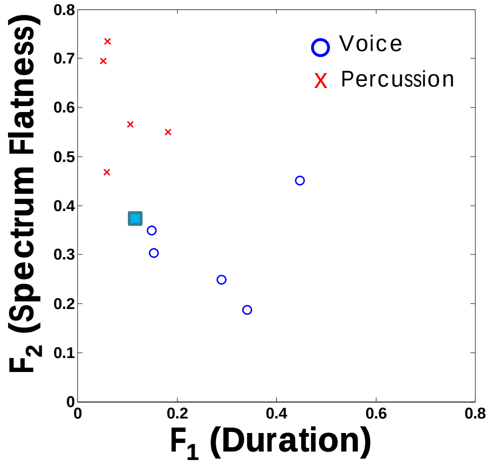
$$p(F_2 = 0.37|V) \propto \frac{1}{\sqrt{2\pi \times 0.0120}} e^{-\frac{(0.37-0.6040)^2}{2\times 0.0120}} \propto 0.3719$$

$$p(F_1 = 0.14|V)\,p(F_2 = 0.37|V) \propto 4.9611 \times 0.3719 = 1.8450$$

Naive Bayes Classifier

$$p(F_1 = 0.14|V)\,p(F_2 = 0.37|V) \propto 6.1600$$
$$p(F_1 = 0.14|P)\,p(F_2 = 0.37|P) \propto 1.8450$$

$$Class(F_1 = 0.14, F_2 = 0.37) = V$$

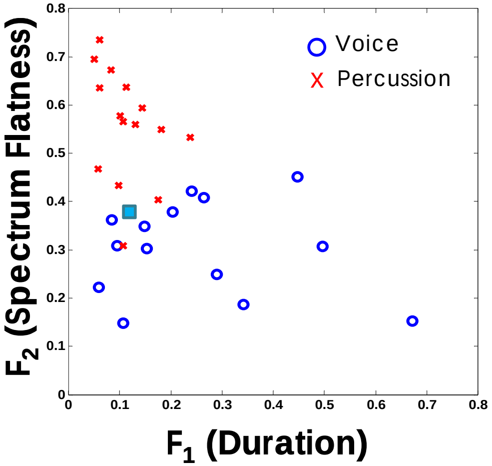


Naive Bayes Classifier

More Training Samples!!!

$$p(F_1 = 0.14|V)\,p(F_2 = 0.37|V) \propto 2.6419$$
$$p(F_1 = 0.14|P)\,p(F_2 = 0.37|P) \propto 6.2181$$

$$Class(F_1 = 0.14, F_2 = 0.37) = P$$



Gaussian Mixture Model

Gaussian Distribution [Single variable]

$$f_G(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$X = \{x_1, x_2, \ldots, x_N\}$$

Training Data

$$\mu \approx \frac{1}{N}\sum_{i=1}^{N} x_i$$

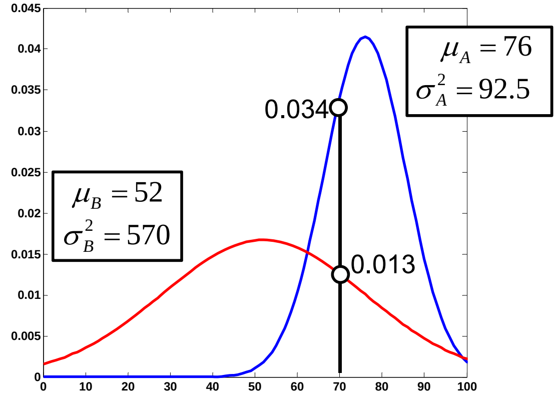
$$\sigma^2 \approx \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2$$

Gaussian Mixture Model

คณะ	ความเกรียน
A	80
A	70
A	90
A	65
A	75
B	80
B	70
B	50
B	20
B	40
?	70

$\Uparrow$

$A \Leftarrow p_A > p_B$



ไม่ต่างอะไรจาก Naive Bayes

Gaussian Mixture Model

Gaussian Mixture Model [Single variable]

$$\sum_{i=1}^{M} c_i = 1$$

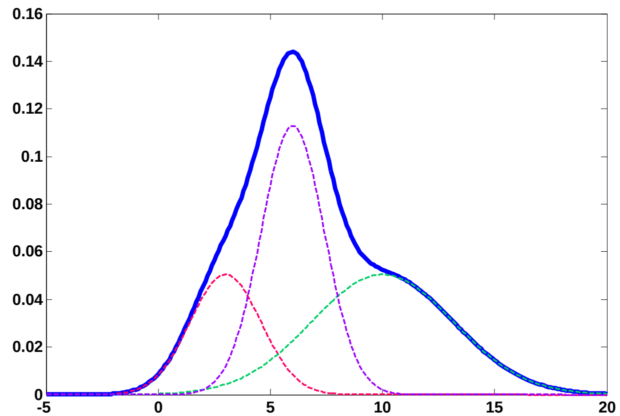
$$f_G(x|\mu_i,\sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}$$

$$\boxed{f_{GMM}(x|\lambda) = \sum_{i=1}^{M} c_i\, f_G(x|\mu_i,\sigma_i)}$$

$$\lambda = \{c_1, c_2, \ldots, c_M, \mu_1, \mu_2, \ldots, \mu_M, \sigma_1, \sigma_2, \ldots, \sigma_M\}$$

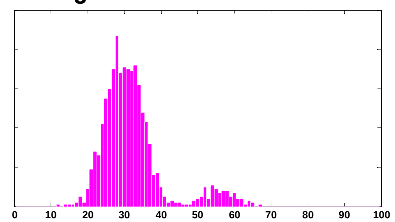
Gaussian Mixture Model

$$f_{GMM}(x) = \frac{0.2}{\sqrt{5\pi}} e^{-\frac{(x-3)^2}{5}} + \frac{0.4}{\sqrt{4\pi}} e^{-\frac{(x-6)^2}{4}} + \frac{0.4}{\sqrt{20\pi}} e^{-\frac{(x-10)^2}{20}}$$



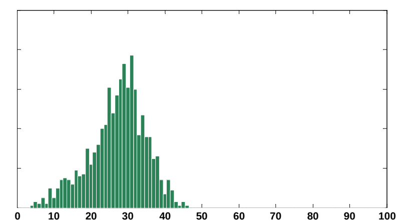
Gaussian Mixture Model

Histogram of Score of Course A



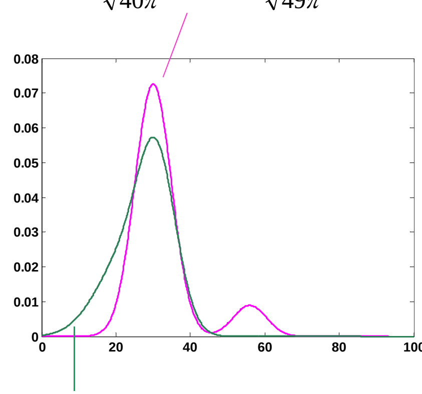
Train GMM for Course A

Histogram of Score of Course B



Train GMM for Course B

$$f_A(x) = \frac{0.1}{\sqrt{40\pi}} e^{-\frac{(x-56)^2}{40}} + \frac{0.9}{\sqrt{49\pi}} e^{-\frac{(x-30)^2}{49}}$$



$$f_B(x) = \frac{0.6}{\sqrt{54\pi}} e^{-\frac{(x-31)^2}{54}} + \frac{0.4}{\sqrt{114\pi}} e^{-\frac{(x-22)^2}{114}}$$

Gaussian Mixture Model

$X = \{x_1, x_2, \ldots, x_N\}$

Training Data

$\lambda = \{c_i, \mu_i, \sigma_i\}$

Training GMM

- Find $\{c_i, \mu_i, \sigma_i\}$ for $i = 1, 2, \ldots, M$
- To maximize likelihood function

$$L(\lambda | X) = \prod_{i=1}^{N} f(x_i | \lambda)$$

- Subject to a constraint $\int_{-\infty}^{\infty} f(x)dx = 1$
- Expectation Maximization (EM) algorithm can be used for solving this optimization problem

Gaussian Mixture Model

For Multiple Variable

คณะ	ความเกรียน	ความขี้เหล้า	ความม่อ
A	80	90	60
A	70	40	70
A	90	60	80
A	65	80	80
A	75	85	70
B	80	75	30
B	70	75	20
B	50	80	40
B	20	65	50
B	40	60	30
?	70	70	50

$\mathbf{x}_1[1]$

$$\mathbf{x}_1 = \begin{bmatrix} 80 \\ 90 \\ 60 \end{bmatrix} \qquad \mathbf{x}_2 = \begin{bmatrix} 70 \\ 40 \\ 70 \end{bmatrix}$$

$\mathbf{x}_2[3]$

$$\mathbf{x}_i[j] = The\ j^{th}\ Feature\ of\ the\ i^{th}\ Sample$$

Gaussian Mixture Model

Multivariate Gaussian Distribution

$$f_G(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^k \det(\boldsymbol{\Sigma})}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

$X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$

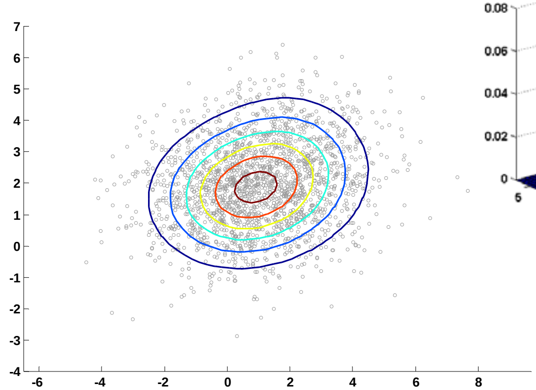
Training Data

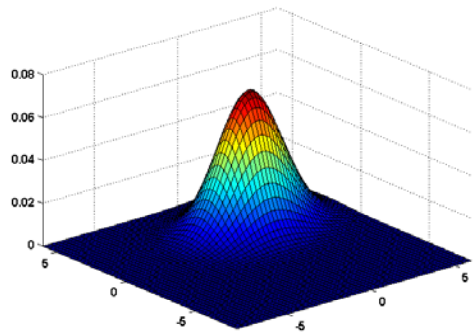
$$\boldsymbol{\mu} \approx \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i$$

$$\boldsymbol{\Sigma}[p,q] \approx \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i[p] - \boldsymbol{\mu}[p])(\mathbf{x}_i[q] - \boldsymbol{\mu}[q])$$

Gaussian Mixture Model

Multivariate Gaussian Distribution





Gaussian Mixture Model

คณะ	เกรียน	ขี้เหล้า	ม่อ
A	80	90	60
A	70	40	70
A	90	60	80
A	65	80	80
A	75	85	70
B	80	75	30
B	70	75	20
B	50	80	40
B	20	65	50
B	40	60	30
?	70	70	50

ความเกรียนเฉลี่ยของกลุ่ม A

$$\boldsymbol{\mu}_A = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i = \frac{1}{5}\times\begin{bmatrix}80+70+90+65+75\\90+40+60+80+85\\60+70+80+80+70\end{bmatrix} = \begin{bmatrix}76\\71\\72\end{bmatrix}$$

$$\boldsymbol{\mu}_B = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i = \frac{1}{5}\times\begin{bmatrix}80+70+50+20+40\\75+75+80+65+60\\30+20+40+50+30\end{bmatrix} = \begin{bmatrix}52\\71\\34\end{bmatrix}$$

ความม่อเฉลี่ยของกลุ่ม B

Gaussian Mixture Model

คณะ	เกรียน	ขี้เหล้า	ม่อ
A	80	90	60
A	70	40	70
A	90	60	80
A	65	80	80
A	75	85	70
B	80	75	30
B	70	75	20
B	50	80	40
B	20	65	50
B	40	60	30
?	70	70	50

Covariance Matrix

$$\boldsymbol{\Sigma}[p,q] = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i[p]-\boldsymbol{\mu}[p])(\mathbf{x}_i[q]-\boldsymbol{\mu}[q])$$

$$\boldsymbol{\Sigma}_A = \begin{bmatrix}92.5 & -1.25 & -2.5\\-1.25 & 430 & -52.5\\-2.5 & -52.5 & 70\end{bmatrix}$$

$$\boldsymbol{\Sigma}_B = \begin{bmatrix}570 & 122.5 & -210\\122.5 & 67.5 & -17.5\\-210 & -17.5 & 130\end{bmatrix}$$

Gaussian Mixture Model

$$\boldsymbol{\Sigma}[p,q] = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i[p]-\boldsymbol{\mu}[p])(\mathbf{x}_i[q]-\boldsymbol{\mu}[q])$$

$$\boldsymbol{\Sigma}_A[1,1] = \frac{1}{5-1}\sum_{i=1}^{5}(\mathbf{x}_i[1]-\boldsymbol{\mu}[1])(\mathbf{x}_i[1]-\boldsymbol{\mu}[1])$$
$$= \frac{(80-76)^2+(70-76)^2+(90-76)^2+(65-76)^2+(75-76)^2}{4}$$

$$\boldsymbol{\Sigma}_A[1,2] = \frac{1}{5-1}\sum_{i=1}^{5}(\mathbf{x}_i[1]-\boldsymbol{\mu}[1])(\mathbf{x}_i[2]-\boldsymbol{\mu}[2])$$
$$= \frac{(80-76)(90-71)+(70-76)(40-71)+(90-76)(60-71)+(65-76)(80-71)+(75-76)(85-71)}{4}$$

Gaussian Mixture Model

$$\mathbf{x} = \begin{bmatrix}70\\70\\50\end{bmatrix} \quad \boldsymbol{\mu}_A = \begin{bmatrix}76\\71\\72\end{bmatrix} \quad \boldsymbol{\Sigma}_A = \begin{bmatrix}92.5 & -1.25 & -2.5\\-1.25 & 430 & -52.5\\-2.5 & -52.5 & 70\end{bmatrix} \quad \det(\boldsymbol{\Sigma}_A) = 2.53\times10^6 \quad \boldsymbol{\Sigma}_A^{-1} = \begin{bmatrix}0.0108 & 0.0001 & 0.0005\\0.0001 & 0.0026 & 0.0019\\0.0005 & 0.0019 & 0.0157\end{bmatrix}$$

$$f(\mathbf{x}|\boldsymbol{\mu}_A,\boldsymbol{\Sigma}_A) = \frac{1}{\sqrt{(2\pi)^k\det(\boldsymbol{\Sigma}_A)}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_A)^T\boldsymbol{\Sigma}_A^{-1}(\mathbf{x}-\boldsymbol{\mu}_A)} = \frac{1}{\sqrt{(2\pi)^3 2.53\times10^6}}e^{-\frac{8.2175}{2}} = 6.5579\times10^{-7}$$

$$(\mathbf{x}-\boldsymbol{\mu}_A)^T\boldsymbol{\Sigma}_A^{-1}(\mathbf{x}-\boldsymbol{\mu}_A) = \left(\begin{bmatrix}70\\70\\50\end{bmatrix}-\begin{bmatrix}76\\71\\72\end{bmatrix}\right)^T\begin{bmatrix}0.0108 & 0.0001 & 0.0005\\0.0001 & 0.0026 & 0.0019\\0.0005 & 0.0019 & 0.0157\end{bmatrix}\left(\begin{bmatrix}70\\70\\50\end{bmatrix}-\begin{bmatrix}76\\71\\72\end{bmatrix}\right)$$
$$= \begin{bmatrix}-6 & -1 & -22\end{bmatrix}\begin{bmatrix}0.0108 & 0.0001 & 0.0005\\0.0001 & 0.0026 & 0.0019\\0.0005 & 0.0019 & 0.0157\end{bmatrix}\begin{bmatrix}-6\\-1\\-22\end{bmatrix}$$
$$= 8.2175$$

Gaussian Mixture Model

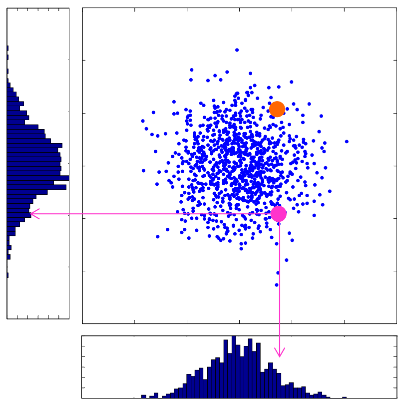
$$\mathbf{x}=\begin{bmatrix}70\\70\\50\end{bmatrix} \quad \boldsymbol{\mu}_B=\begin{bmatrix}52\\71\\34\end{bmatrix} \quad \boldsymbol{\Sigma}_B=\begin{bmatrix}570 & 122.5 & -210\\122.5 & 67.5 & -17.5\\-210 & -17.5 & 130\end{bmatrix} \quad \det(\boldsymbol{\Sigma}_B)=8\times10^5 \quad \boldsymbol{\Sigma}_B^{-1}=\begin{bmatrix}0.011 & -0.015 & 0.015\\-0.015 & 0.038 & -0.020\\0.015 & -0.020 & 0.029\end{bmatrix}$$

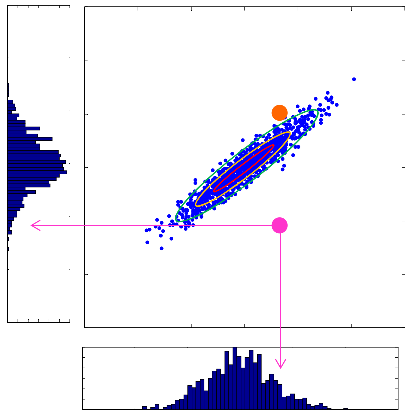
$$f(\mathbf{x}|\boldsymbol{\mu}_B,\boldsymbol{\Sigma}_B)=\frac{1}{\sqrt{(2\pi)^k\det(\boldsymbol{\Sigma}_B)}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_B)^T\boldsymbol{\Sigma}_B^{-1}(\mathbf{x}-\boldsymbol{\mu}_B)}=2.1377\times10^{-9}$$

$$p_A > p_B \Rightarrow Class(70,70,50)=A$$

Gaussian Mixture Model

$$P(F_1,F_2)=P(F_1)P(F_2)$$





Naive Bayes

Gaussian Mixture Model

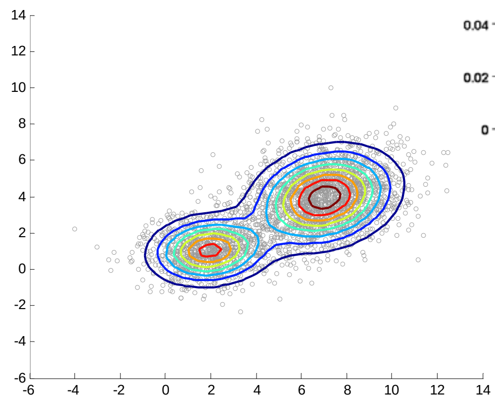
Gaussian Mixture Model [multiple variables]

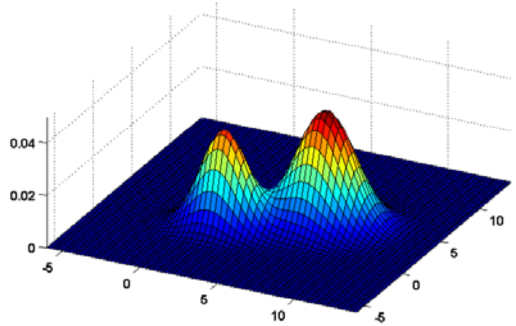
$$\sum_{i=1}^{M}c_i=1 \qquad f_G(\mathbf{x}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)=\frac{1}{\sqrt{(2\pi)^k\det(\boldsymbol{\Sigma}_i)}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)}$$

$$\boxed{f_{GMM}(\mathbf{x}|\lambda)=\sum_{i=1}^{M}c_i\,f_G(\mathbf{x}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)}$$

$$\lambda=\left\{c_1,c_2,\ldots,c_M,\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\ldots,\boldsymbol{\mu}_M,\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2,\ldots,\boldsymbol{\Sigma}_M\right\}$$

Gaussian Mixture Model





Gaussian Mixture Model

Training Data $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$

$\Downarrow$ Training GMM

$$\lambda = \{c_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\} \text{for } i = 1,2,3,\ldots,M$$

$\Downarrow$

$$\mathbf{x} \Rightarrow f_{GMM}(\mathbf{x}|\lambda) = \sum_{i=1}^{M} c_i f_G(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \Rightarrow f(\mathbf{x})$$

Gaussian Mixture Model

Expectation Maximization Algorithm

[EM Algorithm]

Initial Guess of $\{c_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$ for $i = 1,2,3,\ldots,M$

$\Downarrow$

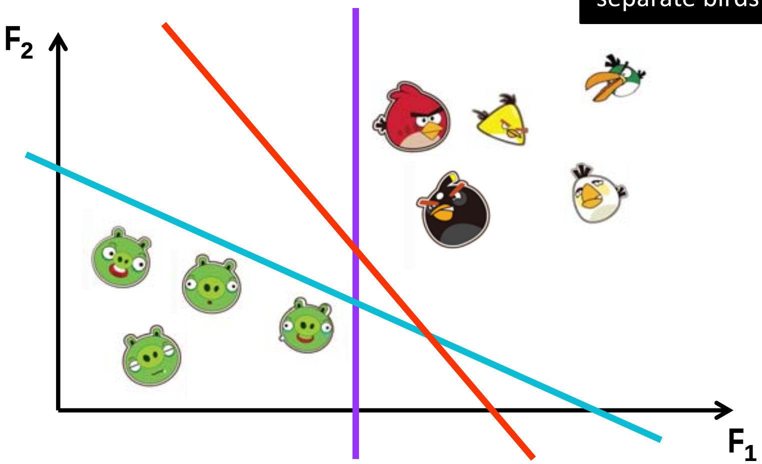
$$p(i|\mathbf{x}_k, \lambda) = \frac{c_i f_G(\mathbf{x}_k|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{f_{GMM}(\mathbf{x}_k|\lambda)} \Leftarrow c_i^{new} = \frac{1}{N}\sum_{k=1}^{N} p(i|\mathbf{x}_k, \lambda)$$

$\Downarrow$ $\Uparrow$

$$\boldsymbol{\mu}_i^{new} = \frac{\sum_{k=1}^{N} p(i|\mathbf{x}_k, \lambda)\mathbf{x}_k}{\sum_{k=1}^{N} p(i|\mathbf{x}_k, \lambda)} \Rightarrow \boldsymbol{\Sigma}_i^{new} = \frac{\sum_{k=1}^{N} p(i|\mathbf{x}_k, \lambda)(\mathbf{x}_k - \boldsymbol{\mu}_i)^T(\mathbf{x}_k - \boldsymbol{\mu}_i)}{\sum_{k=1}^{N} p(i|\mathbf{x}_k, \lambda)}$$

Support Vector Machine

Which line is the best for separate birds and pigs?



Support Vector Machine

Training Set

$$\{(\mathbf{x}_1, c_1), (\mathbf{x}_2, c_2), (\mathbf{x}_3, c_3) \cdots (\mathbf{x}_N, c_N)\}$$

$$c_i \in \{-1,1\}$$

$\Downarrow$

Hyperplane

Feature Vector of Unknown Class

$$\mathbf{y}$$

$\Downarrow$

Hyperplane

$\Downarrow$

$$Class(\mathbf{y})$$

Support Vector Machine

Hyperplane

$$w_1x_1 + w_2x_2 + \ldots + w_Nx_N + b = 0$$

$$\Downarrow \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}$$

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

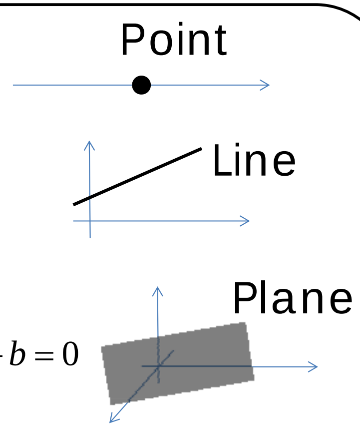
Support Vector Machine

Hyperplane

$N = 1;\ w_1x_1 + b = 0$

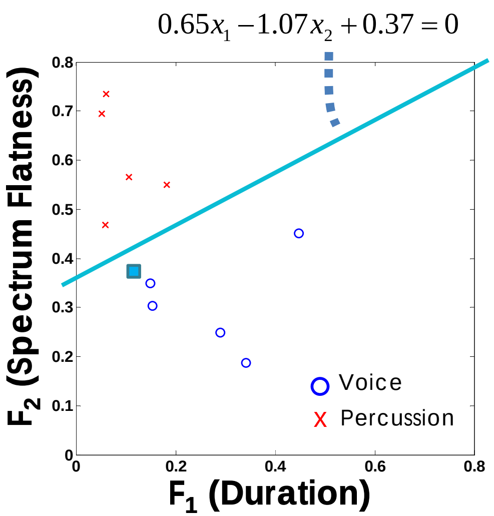
$N = 2;\ w_1x_1 + w_2x_2 + b = 0$

$N = 3;\ w_1x_1 + w_2x_2 + w_3x_3 + b = 0$



Support Vector Machine

Class	F_1	F_2
V	0.34	0.19
V	0.15	0.30
V	0.45	0.45
V	0.28	0.25
V	0.13	0.35
P	0.11	0.57
P	0.06	0.74
P	0.06	0.47
P	0.18	0.55
P	0.05	0.69
?	0.14	0.37



Support Vector Machine

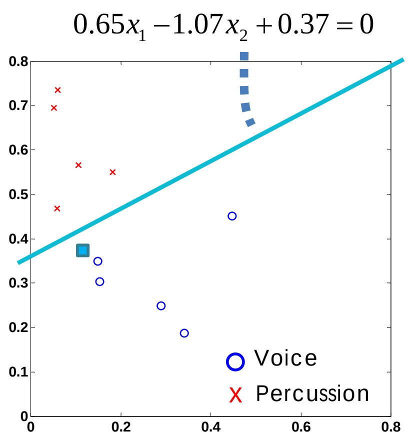
Classify

$$class(\mathbf{y}) = \begin{cases} 1; & \mathbf{w}^T\mathbf{y} + b > 0 \\ -1; & \mathbf{w}^T\mathbf{y} + b < 0 \end{cases}$$

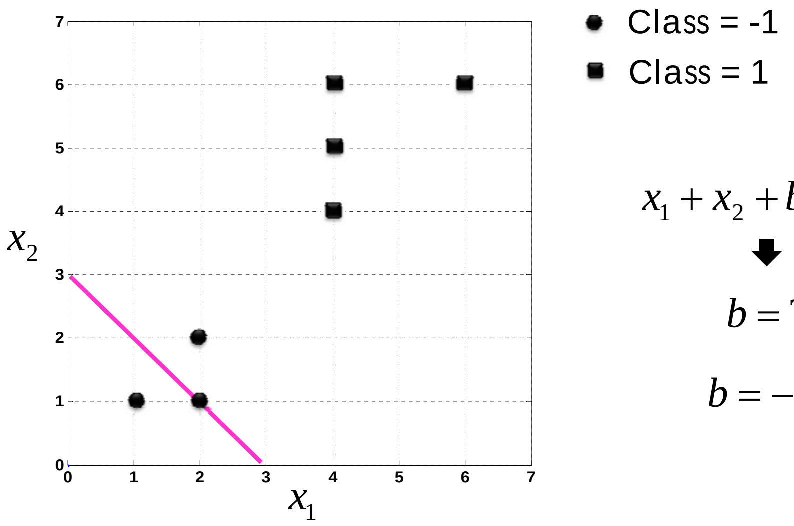
$\mathbf{y} = \begin{bmatrix} 0.14 \\ 0.37 \end{bmatrix}$ $\Downarrow$

$0.65(0.14) - 1.07(0.37) + 0.37 = 0.0651$

$\Uparrow$ $class(\mathbf{y}) = 1$

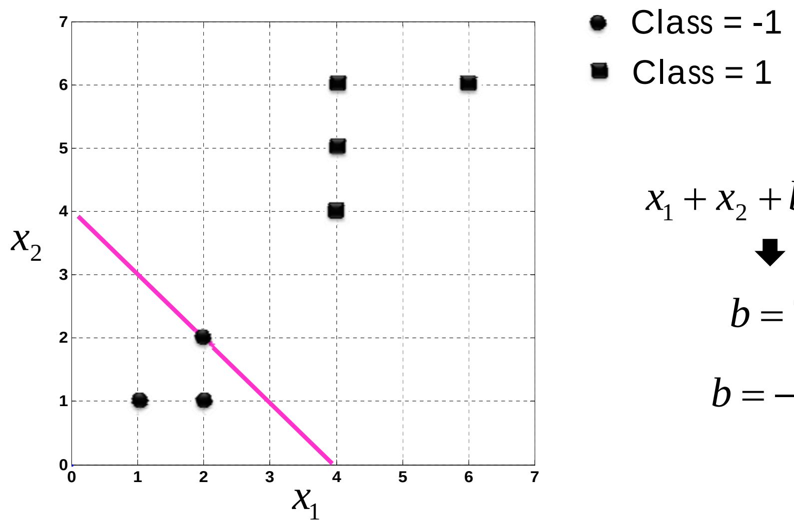


Support Vector Machine



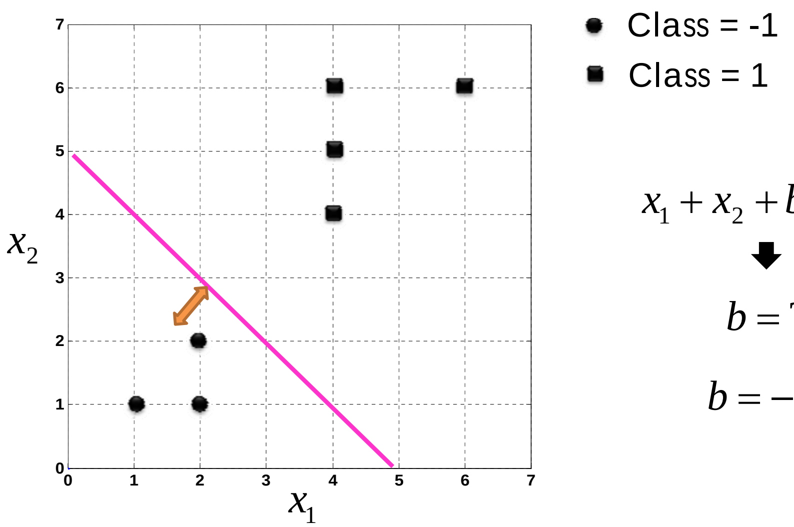
$$x_1 + x_2 + b = 0$$
$$\Downarrow$$
$$b = ?$$
$$b = -3?$$

Support Vector Machine



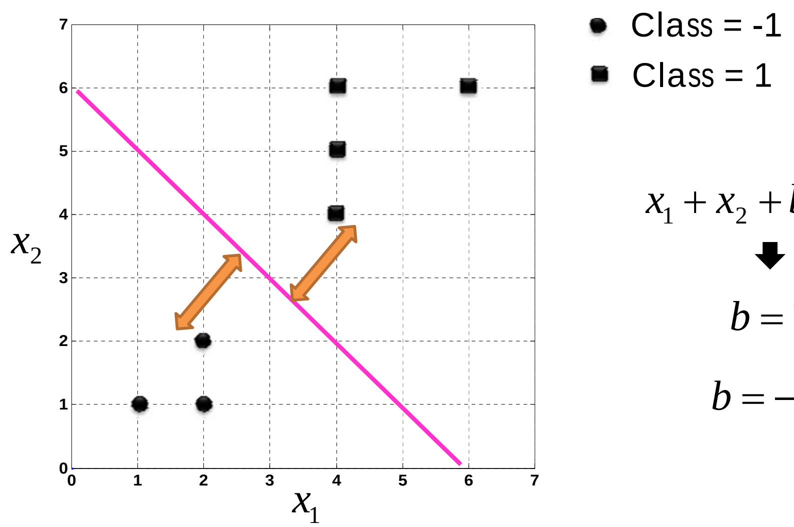
$$x_1 + x_2 + b = 0$$
$$\Downarrow$$
$$b = ?$$
$$b = -4?$$

Support Vector Machine



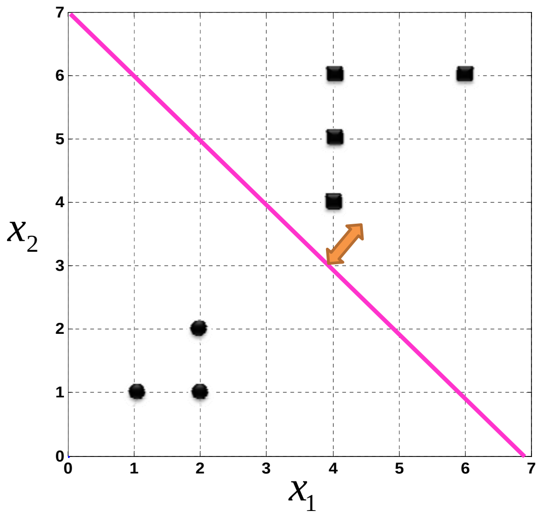
$$x_1 + x_2 + b = 0$$
$$\Downarrow$$
$$b = ?$$
$$b = -5?$$

Support Vector Machine



$$x_1 + x_2 + b = 0$$
$$\Downarrow$$
$$b = ?$$
$$b = -6?$$

Support Vector Machine



- Class = -1
- Class = 1

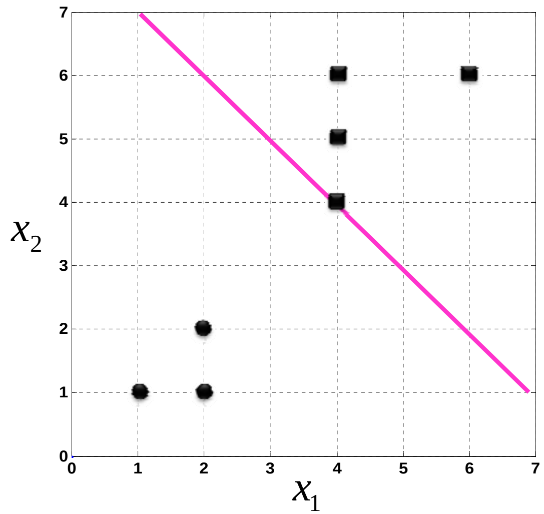
$$x_1 + x_2 + b = 0$$

$$\Downarrow$$

$$b = ?$$

$$b = -7\,?$$

Support Vector Machine



- Class = -1
- Class = 1

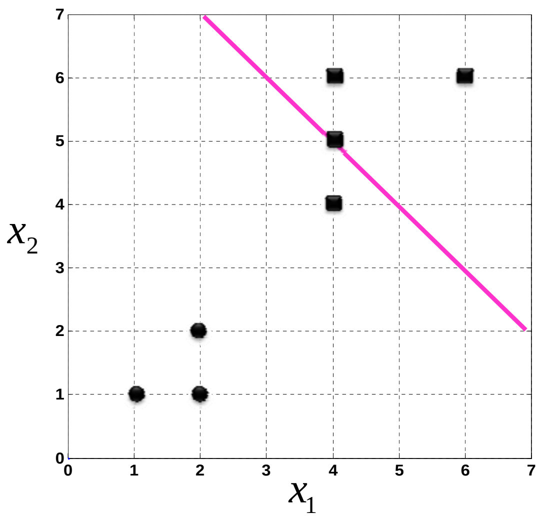
$$x_1 + x_2 + b = 0$$

$$\Downarrow$$

$$b = ?$$

$$b = -8\,?$$

Support Vector Machine



- Class = -1
- Class = 1

$$x_1 + x_2 + b = 0$$

$$\Downarrow$$

$$b = ?$$

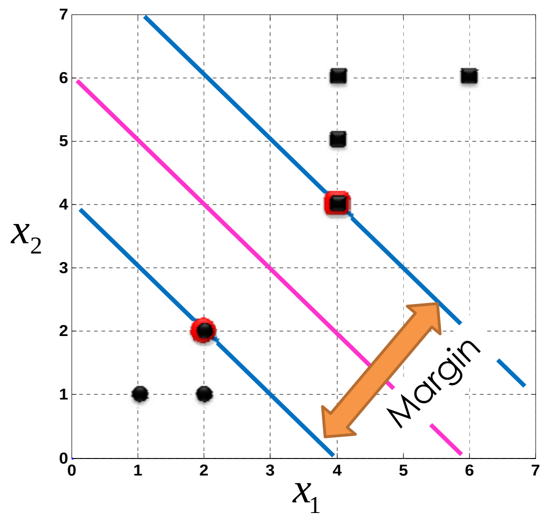
$$b = -9\,?$$

Support Vector Machine

$$x_1 + x_2 - 6 = 0$$

is the best

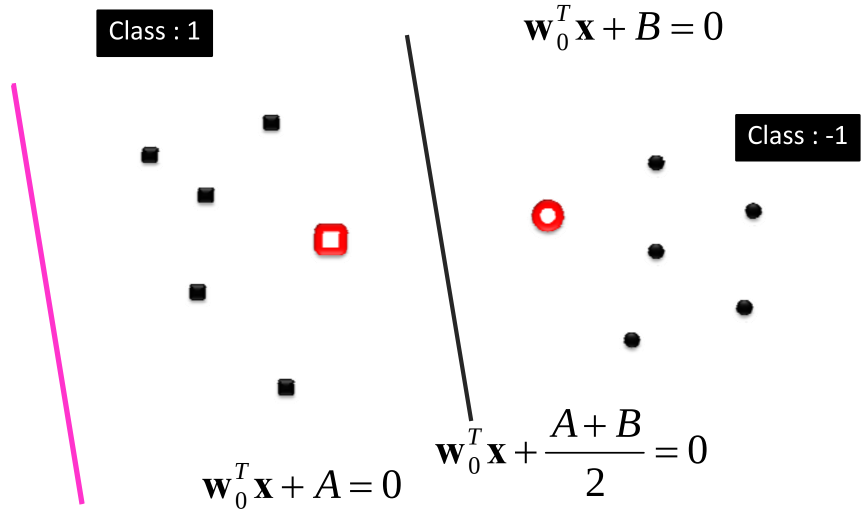
$$\boxed{\frac{(x_1 + x_2 - 4 = 0) + (x_1 + x_2 - 8 = 0)}{2}}$$



○ □

Support Vector

Support Vector Machine



Support Vector Machine

Class : 1

$$\mathbf{w}_0^T\mathbf{x} + A = 0 \Rightarrow \mathbf{w}_0^T\mathbf{x} + \frac{A+B}{2} = \frac{B-A}{2} \Rightarrow \mathbf{w}^T\mathbf{x} + b = 1$$

Class : -1

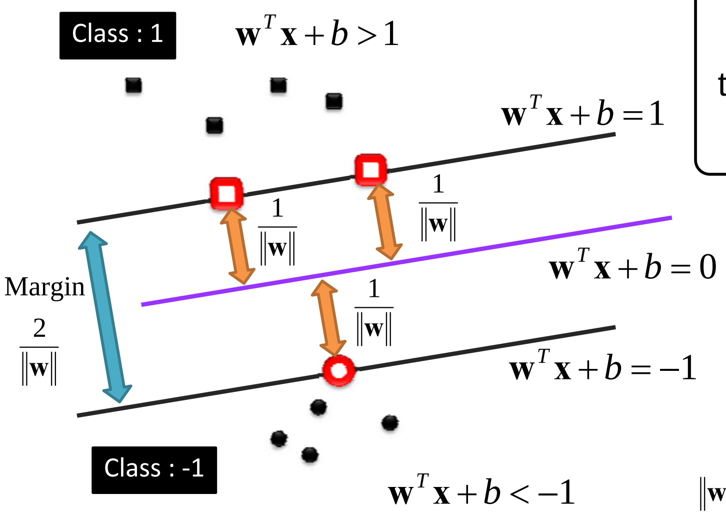
$$\mathbf{w}_0^T\mathbf{x} + B = 0 \Rightarrow \mathbf{w}_0^T\mathbf{x} + \frac{A+B}{2} = \frac{A-B}{2} \Rightarrow \mathbf{w}^T\mathbf{x} + b = -1$$

Middle

$$\mathbf{w}_0^T\mathbf{x} + \frac{A+B}{2} = 0 \Rightarrow \mathbf{w}^T\mathbf{x} + b = 0$$

$$\boxed{\mathbf{w} = \frac{2}{B-A}\mathbf{w}_0 \quad \Big| \quad b = \frac{A+B}{B-A}}$$

Support Vector Machine



Target

Find $\mathbf{w}, b$ that maximize margin

$$\|\mathbf{w}\| = \sqrt{w_1^2 + w_2^2 + \ldots + w_M^2}$$

Support Vector Machine

$$\{(\mathbf{x}_1, c_1), (\mathbf{x}_2, c_2), (\mathbf{x}_3, c_3) \cdots (\mathbf{x}_N, c_N)\}$$
$$c_i \in \{-1, 1\}$$

Training Data

$$\Downarrow$$

$$\{\mathbf{w}, b\}$$

Training SVM

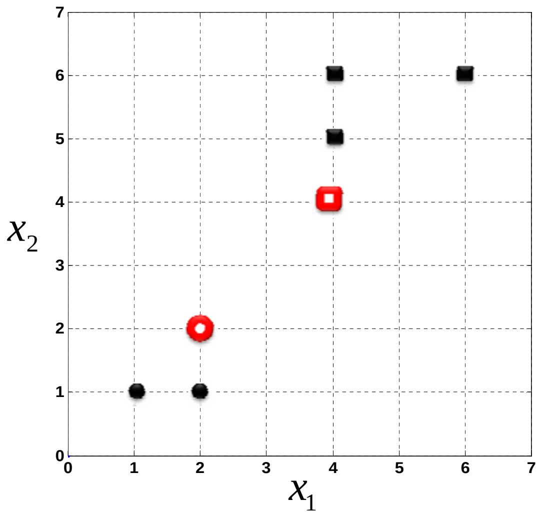
- Find $\{\mathbf{w}, b\}$
- To minimize $\|\mathbf{w}\|$
- Subject to a constraint

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \text{ for } \forall i\{c_i = 1\}$$
$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \text{ for } \forall i\{c_i = -1\}$$

$$\Downarrow$$

$$c_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \text{ for every } i$$

Support Vector Machine



- Class = -1
- Class = 1

$$\left\{\left(\begin{bmatrix}2\\2\end{bmatrix},-1\right),\left(\begin{bmatrix}4\\4\end{bmatrix},1\right)\right\} = \text{Support Vector}$$

$$\Downarrow$$

$$\mathbf{w}, b = ?$$

Support Vector Machine

$$\left\{\left(\begin{bmatrix}2\\2\end{bmatrix},-1\right),\left(\begin{bmatrix}4\\4\end{bmatrix},1\right)\right\} = \text{Support Vector}$$

$$\Downarrow$$

$$\mathbf{w}^T\begin{bmatrix}2\\2\end{bmatrix}+b=-1$$

$$\mathbf{w}^T\begin{bmatrix}4\\4\end{bmatrix}+b=1$$

$$\Rightarrow$$

- Find $\{\mathbf{w}, b\}$
- To minimize $\|\mathbf{w}\| = \sqrt{w_1^2 + w_2^2}$
- Subject to a constraint

$$2w_1 + 2w_2 + b = -1$$
$$4w_1 + 4w_2 + b = 1$$

$$\Downarrow$$

$$w_1 + w_2 = 1$$

Support Vector Machine

Minimize $\|\mathbf{w}\| = \sqrt{w_1^2 + w_2^2}$ when $w_1 + w_2 = 1$

$$\|\mathbf{w}\|^2 = w_1^2 + w_2^2 = w_1^2 + (1-w_1)^2$$

$$\frac{\partial\|\mathbf{w}\|^2}{\partial w_1} = 2w_1 + 2(1-w_1)$$

$$0 = 2w_1 + 2(1-w_1)$$

$$w_1 = 0.5$$

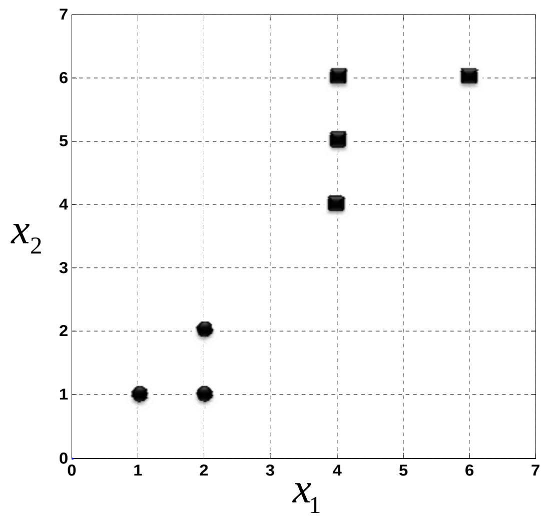
$$w_2 = 1 - w_1 = 0.5$$

$$2w_1 + 2w_2 + b = -1$$
$$b = -1 - 2w_1 - 2w_2 = -3$$

Max Margin hyperplane

$$\boxed{0.5x_1 + 0.5x_2 - 3 = 0}$$

Support Vector Machine



- Class = -1
- Class = 1

$$w_1 + w_2 + b \le -1$$
$$2w_1 + w_2 + b \le -1$$
$$2w_1 + 2w_2 + b \le -1$$

$$4w_1 + 4w_2 + b \ge 1$$
$$4w_1 + 5w_2 + b \ge 1$$
$$4w_1 + 6w_2 + b \ge 1$$
$$6w_1 + 6w_2 + b \ge 1$$

$$\Downarrow$$

$$\mathbf{w}, b = ?$$

Support Vector Machine

- Find $\mathbf{x}$
- To maximize $f(\mathbf{x})$
- Subject to a constraint $g(\mathbf{x}) = K$

Lagrange Function

$$\Lambda(\mathbf{x},\alpha) = f(\mathbf{x}) + \alpha\big(g(\mathbf{x}) - K\big)$$

α : Lagrange Multiplier

- Find critical points where $\frac{\partial \Lambda}{\partial x_i} = 0, \frac{\partial \Lambda}{\partial \alpha} = 0$

Support Vector Machine

Training SVM

Find $\{\mathbf{w}, b\}$ to minimize $\|\mathbf{w}\|$ subject to $c_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1; \forall i$

Lagrange Function

$$\Lambda(\mathbf{w},b,\boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N}\alpha_i\left[c_i(\mathbf{w}^T\mathbf{x}_i + b) - 1\right]$$

Find $\{\mathbf{w}, b\}$ to minimize Λ subject to $\alpha_i \geq 0; \forall i$

Support Vector Machine

$$\Lambda(\mathbf{w},b,\boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N}\alpha_i\left[c_i(\mathbf{w}^T\mathbf{x}_i + b) - 1\right]$$

Find $\{\mathbf{w}, b\}$ to minimize Λ subject to $\alpha_i \geq 0; \forall i$

Lagrange Duality

Dual form

$$\Lambda_D(\boldsymbol{\alpha}) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j c_i c_j \mathbf{x}_i^T\mathbf{x}_j$$

Find $\boldsymbol{\alpha}$ to maximize Λ_D subject to $\alpha_i \geq 0; \forall i$

Support Vector Machine

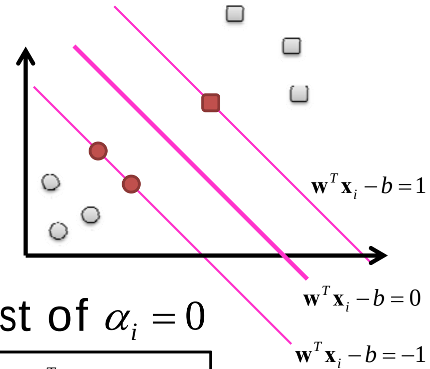
$\{(\mathbf{x}_1,c_1),(\mathbf{x}_2,c_2),(\mathbf{x}_3,c_3)\cdots(\mathbf{x}_N,c_N)\}$

$c_i \in \{-1,1\}$

Training Data

Lagrange Multiplier

$\{\alpha_1, \alpha_2, \ldots, \alpha_N\}$



- Most of $\alpha_i = 0$

  $c_i(\mathbf{w}^T\mathbf{x}_i + b) > 1$

- $\alpha_i > 0 \Rightarrow \mathbf{x}_i =$ **Support Vector**

  $c_i(\mathbf{w}^T\mathbf{x}_i + b) = 1$

Support Vector Machine

Lagrange Multiplier
$\{\alpha_1, \alpha_2, ..., \alpha_N\}$

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i c_i \mathbf{x}_i = \sum_{i \in \substack{\text{Support} \\ \text{Vector}}} \alpha_i c_i \mathbf{x}_i$$

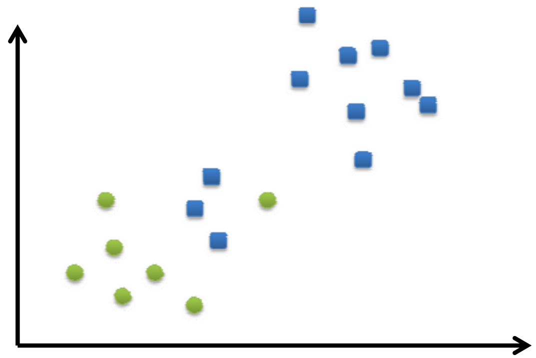
$$b = \mathbf{w}^T \mathbf{x}_i - c_i \; ; i \in \text{Support Vector}$$

Classify

$$class(\mathbf{y}) = sign(\mathbf{w}^T \mathbf{y} + b) = sign\Big(\sum_{i \in \substack{\text{Support} \\ \text{Vector}}} \alpha_i c_i \mathbf{x}_i^T \mathbf{y} + b \Big)$$

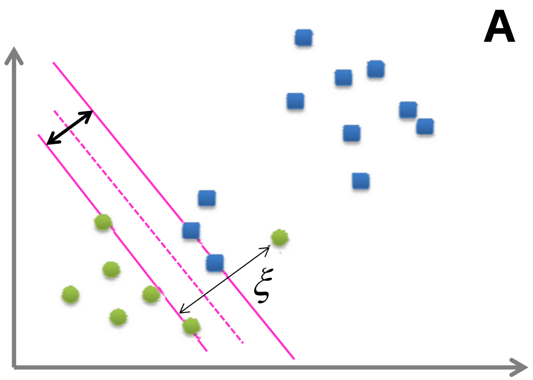
Support Vector Machine

Soft Margin



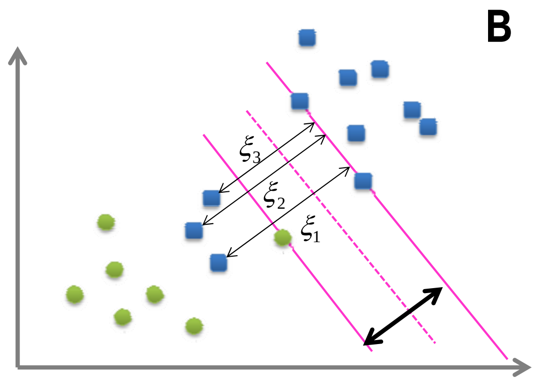
Support Vector Machine

Soft Margin



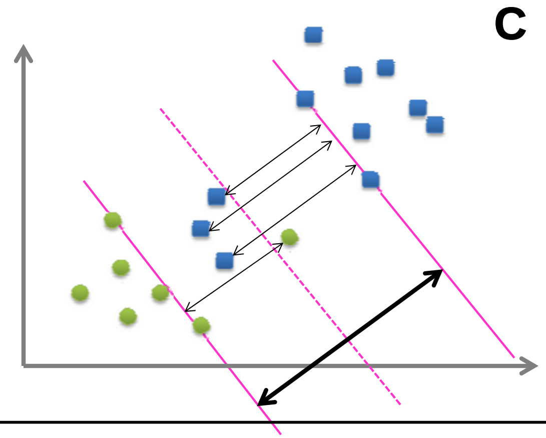
Support Vector Machine

Soft Margin



Support Vector Machine

Soft Margin



Support Vector Machine

Soft Margin

- Find $\{\mathbf{w}, b\}$
- To minimize

$$\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$

- Subject to a constraint

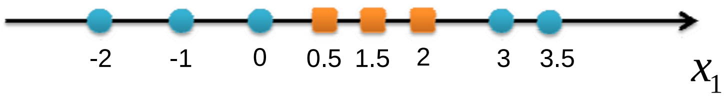
$$c_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \text{ for every } i$$

- Classify

$$class(\mathbf{y}) = sign(\mathbf{w}^T\mathbf{y} + b)$$

Support Vector Machine

Non Linear SVM

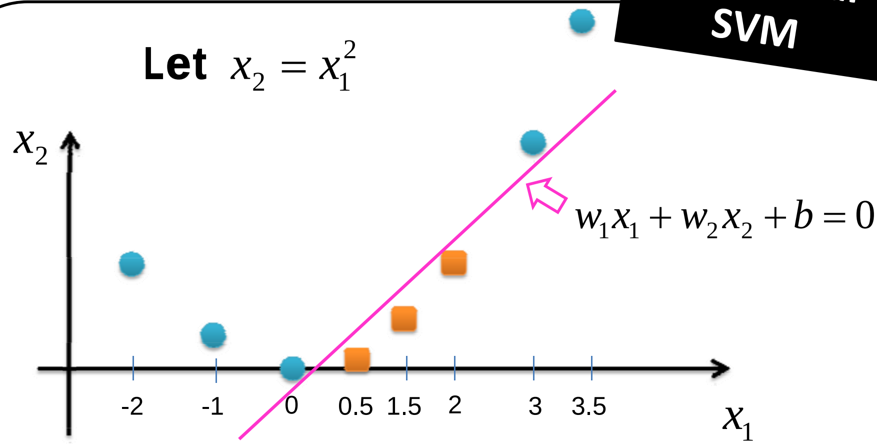


Can not separate by one hyperplane

Support Vector Machine

Non Linear SVM

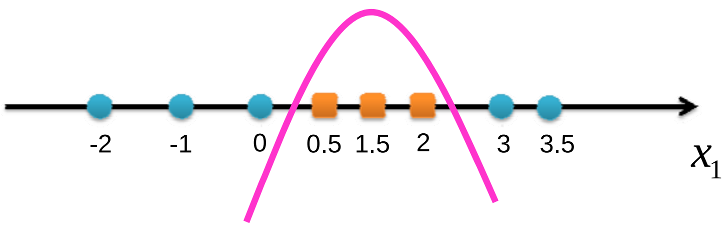
Let $x_2 = x_1^2$



Support Vector Machine

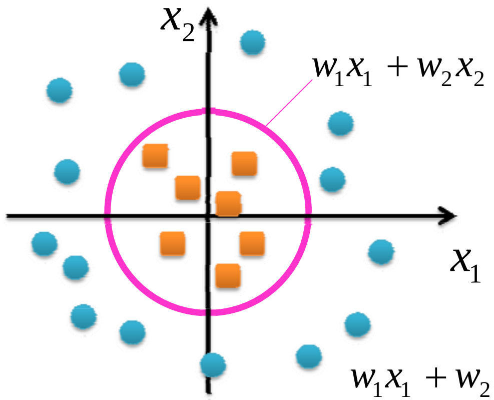
Non Linear SVM

$$w_1x_1 + w_2x_2 + b = 0 \Rightarrow w_1x_1 + w_2x_1^2 + b = 0$$



Support Vector Machine

Non Linear SVM



$$w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2 + b = 0$$

$$x_3 = x_1^2$$
$$x_4 = x_2^2$$

$$w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b = 0$$

Support Vector Machine

Non Linear SVM

Mapping Function

$$\mathbf{x} \Rightarrow \Phi(\mathbf{x})$$

$$x \Rightarrow \begin{bmatrix} x \\ x^2 \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_1^2 \\ x_2^2 \end{bmatrix}$$

Support Vector Machine

Non Linear SVM

- Find $\{\mathbf{w}, b\}$
- To minimize

$$\frac{1}{2}\|\mathbf{w}\|^2$$

- Subject to a constraint

$$c_i(\mathbf{w}^T\Phi(\mathbf{x}_i) + b) \geq 1 \text{ for every } i$$

- Classify

$$class(\mathbf{y}) = sign(\mathbf{w}^T\Phi(\mathbf{y}) + b)$$

$$= sign\Big(\sum_{i\in \text{Support Vector}} \alpha_i c_i \Phi(\mathbf{x}_i)^T \Phi(\mathbf{y}) + b\Big)$$

$$= sign\Big(\sum_{i\in \text{Support Vector}} \alpha_i c_i \underbrace{\kappa(\mathbf{x}_i, \mathbf{y})}_{\text{Kernel Function}} + b\Big)$$

Support Vector Machine

Non Linear SVM

Linear Kernel $\kappa(\mathbf{x},\mathbf{y}) = \mathbf{x}^T\mathbf{y}$

Polynomial Kernel $\kappa(\mathbf{x},\mathbf{y}) = \left(\mathbf{x}^T\mathbf{y}\right)^d$

Radial Basis Kernel $\kappa(\mathbf{x},\mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}}$

Support Vector Machine

Non Linear SVM

Ex. 2nd Order Polynomial Kernel

$$\Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}$$

$$\Phi(\mathbf{x})^T\Phi(\mathbf{y}) = \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}^T \begin{bmatrix} y_1^2 \\ \sqrt{2}y_1y_2 \\ y_2^2 \end{bmatrix} = x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2$$

Support Vector Machine

Non Linear SVM

Ex. 2nd Order Polynomial Kernel

$$\kappa(\mathbf{x},\mathbf{y}) = \left(\mathbf{x}^T\mathbf{y}\right)^d = \left(\mathbf{x}^T\mathbf{y}\right)^2$$

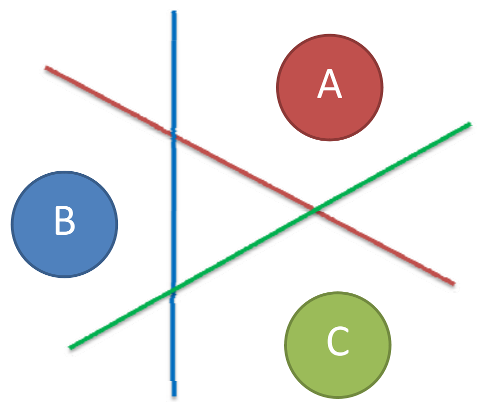
$$= \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right)^2$$

$$= \left(x_1y_1 + x_2y_2\right)^2 = x_1^2y_1^2 + x_1y_1x_2y_2 + x_2^2y_2^2$$

Support Vector Machine

One-Against-All

Output = classifier with the highest output function

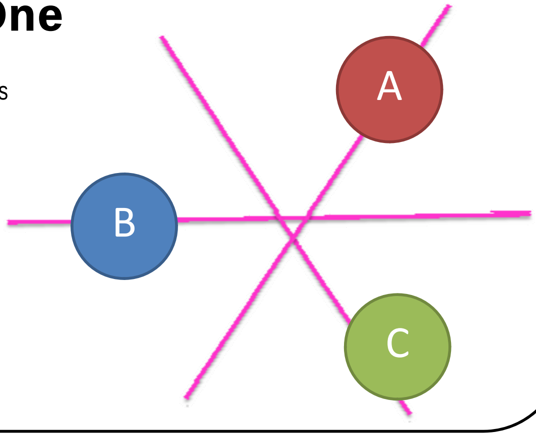


Multi-Class SVM

Support Vector Machine

One-Against-One

Output = maximum votes form classifier



Multi-Class SVM