Machine Learning

Lect 8

Dr. Prerana
Mukherjee

SVM : Support Vector Machines : hyperplane based classifiers

Vapnik Chervonenkis

[VC dimension]

⇓

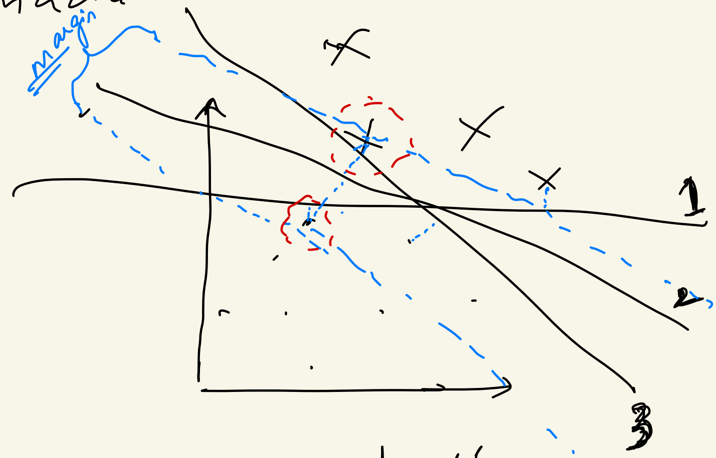
complexity of machine

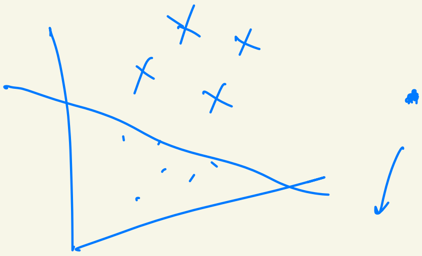
Twin SVM

MCM: Minimal complexity Machine

$\simeq O(n)$

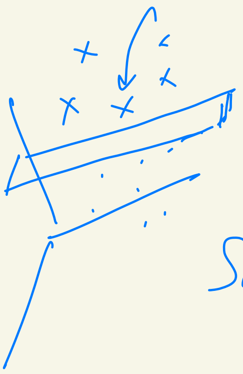
$= o(n)$





1, 2, 3 valid hyperplanes

1, 2, 3 is a good hyperplane?

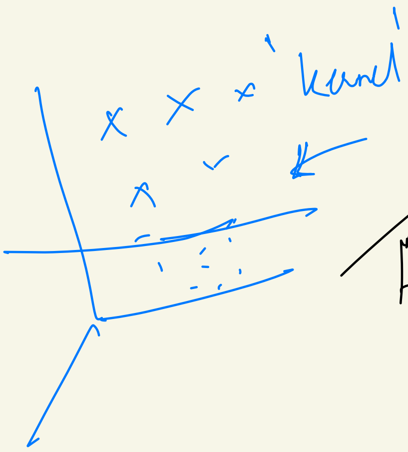


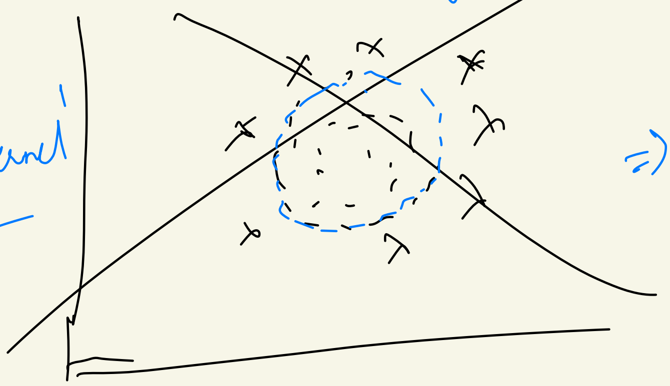
Solution: Find that hyperplane which has maximum margin!

Support vectors

"non linear SVM"

$z^2 = x^2 + y^2$





low dimensional space
is non linearly separable
Mercer's condition
kernel trick

high dimensional space
linearly separable

line → plane

Manifold learning ⇐

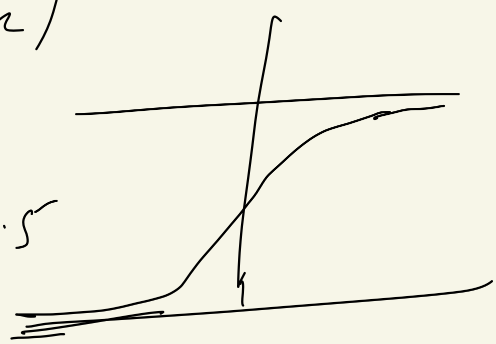
homogeneous coordinate system

$x, y \longrightarrow w \qquad \frac{x}{w}, \frac{y}{w}, w$

Logistic Regression:

$$P(y=1/x) = h(x) = \sigma(\beta^T x)$$

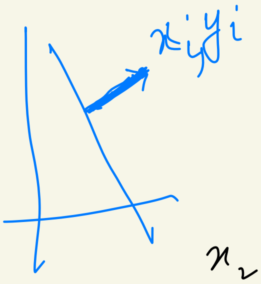
Predict 1 when $h(x) > 0.5$
0 when $h(x) < 0.5$

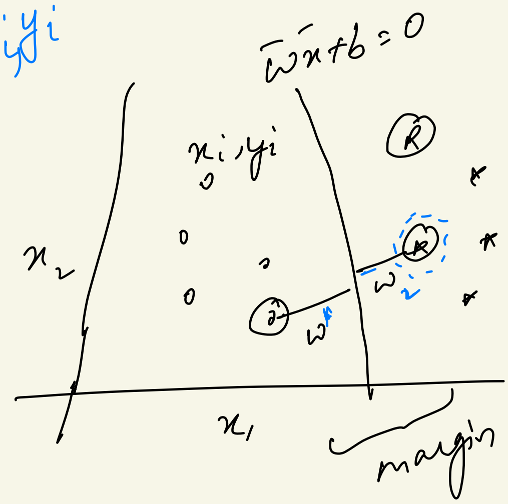


less confidence than points farther from decision boundary

Large no of features
use Bayesian Learning
we look at all possible classifiers
weigh them by their

aposterior probability & combined to get the answer
It is not computationally efficient.





$w_1 = w_2 = w$

$w_1 x_1 + w_2 x_2 + b = 0$

$2w$ = Margin

Goal: margin width is highest & Classifier has highest generalization ability

Points closest to margin: Support vectors

Distance of closest +ve = Distance of closest -ve

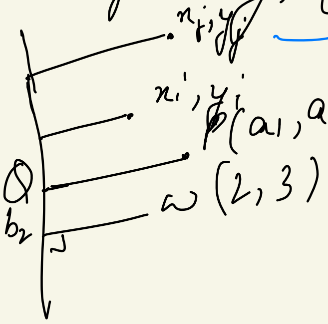
minimum of 2 support vectors

Functional margin: dist of x_i, y_i wrt w, b from decision boundary

$$\gamma^i = y_i(\underbrace{w^T x_i + b}) = \text{Normal to the line}$$

Larger functional margin more confidence to classify

and also depends on coefficients of line (scaling factors)



$S = \{(x_1, y_1) \ldots (x_m, y_m)\}$

$$\boxed{\gamma = \min_{i=1}^{m} \gamma^i}$$

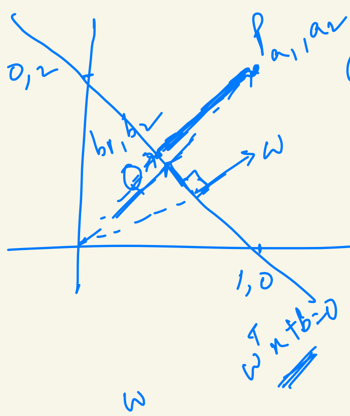
Geometric margin: to get rid of scaling factors

unit vector: $\frac{w}{\|w\|}$

$$= \left(\frac{2}{\sqrt{3}}, \frac{3}{\sqrt{3}}\right)$$

$p = Q + \gamma \frac{w}{\|w\|}$

Line: $2x + 3y + 1 = 0$



$y = -2x + 2$

$2x + y - 2 = 0$

$\underbrace{\begin{bmatrix} 2 & 1 \end{bmatrix}}_{w} \begin{bmatrix} x \\ y \end{bmatrix} \underbrace{-2}_{b} = 0$

$\frac{w}{\|w\|} \left[\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right]$

$$P = Q + \gamma \frac{w}{\|w\|}$$

$$(a_1, a_2) = (b_1, b_2) + \gamma \frac{w}{\|w\|}$$

$$w^T\left((a_1, a_2) - \gamma \frac{w}{\|w\|}\right) + b = 0$$

$$\gamma = \frac{w^T(a_1, a_2) + b}{\|w\|}$$

$\gamma = \min_{i=1}^{m} \gamma_i \rightarrow GM$

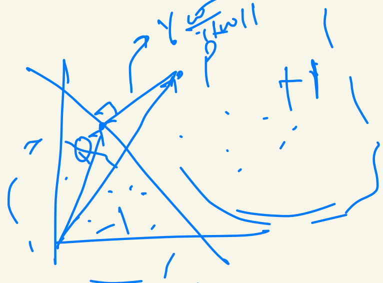
$y\,(w^T x + b)$

$$= \frac{w^T}{\|w\|}(a_1, a_2) + \frac{b}{\|w\|}$$

$$\gamma = y \left(\frac{w^T}{\|w\|}(a_1, a_2) + \frac{b}{\|w\|} \right)$$

Let $\|w\| = 1$ (after normalization)

$x_i, y_i \rightarrow$ $\underset{(GM)}{\gamma} = \underset{\text{class label}}{y}\left(w^T(a_1, a_2) + b\right)$



$P = Q + \gamma \frac{w}{\|w\|}$

$w^T w = \|w\|^2$

$w^T P = w^T Q + \gamma \frac{w^T w}{\|w\|}$

$w^T P = w^T Q + \gamma \|w\|$

Maximize Margin width

→ Maximize $\frac{\gamma}{\|w\|}$ subject to

→ $y_i(w^T x_i + b) \geq \gamma$ for $i = 1, 2 \ldots m$

→ Scale so that $\gamma = 1$ new geometric margin

→ Maximizing $\frac{1}{\|w\|}$ is same as minimizing $\|w\|$ $= w \cdot w$

→ minimize $w \cdot w$ subject to constraints

for all x_i, y_i, $i = 1, \ldots m$:

$$\begin{cases} w^T x_i + b \geq 1 & \text{if } y_i = 1 \\ w^T x_i + b \leq -1 & \text{if } y_i = -1 \end{cases}$$

if (w, b) is decision boundary

$\frac{\gamma}{\|w\|}$ is the geometric margin

Max

Large Margin Linear Classifier

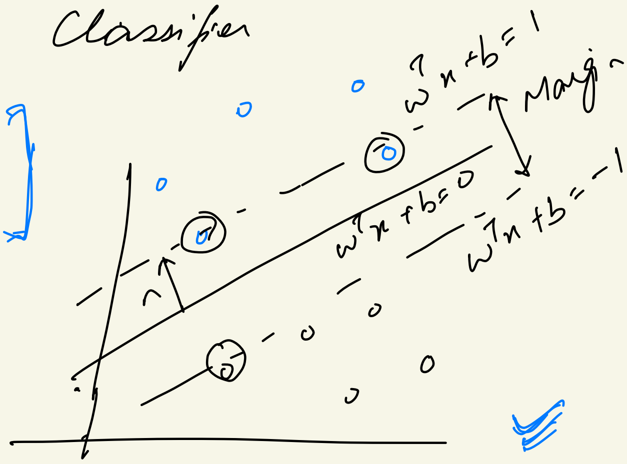
Formulation:

Primal form

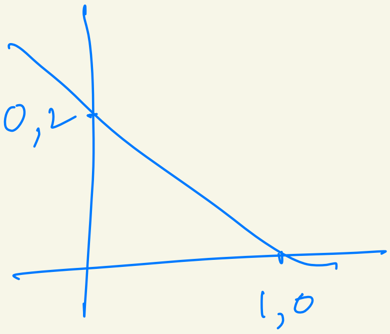
$$\text{minimize} \quad \frac{1}{2} \|w\|^2$$

such that

$$y_i (w^T x_i + b) \geq 1$$



- Optimization problem with convex quadratic objective & m inequality linear constraints
- Can be solved using QP
- Lagrange duality to get the optimization problem's dual form.
- Allow us to use kernels to get optimal margin classifiers to work efficiently in very high dimensional spaces.
- Allow us to derive an efficient algorithm for solving the above optimization problem that will typically do much better than generic QP software.



$$x + \frac{y}{2} = 1$$

$$2x + y = 2$$

$$2x + y - 2 = 0$$

$$w = \begin{bmatrix} 2 & 1 \end{bmatrix}$$

$$\frac{w}{\|w\|} = \begin{bmatrix} \frac{2}{\sqrt{5}}, & \frac{1}{\sqrt{5}} \end{bmatrix}$$