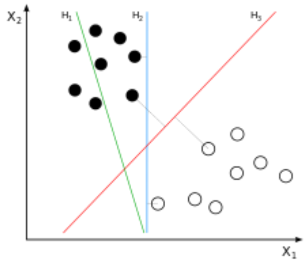
1. 背景：

    1.1 最早是由 Vladimir N. Vapnik 和 Alexey Ya. Chervonenkis 在1963年提出
    1.2 目前的版本(soft margin)是由Corinna Cortes 和 Vapnik在1993年提出，并在1995年发表
    1.3 深度学习（2012）出现之前，SVM被认为机器学习中近十几年来最成功，表现最好的算法

2. 机器学习的一般框架：
    训练集 => 提取特征向量 => 结合一定的算法（分类器：比如决策树，KNN）=>得到结果

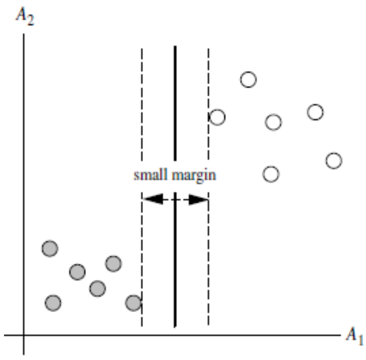
3. 介绍：

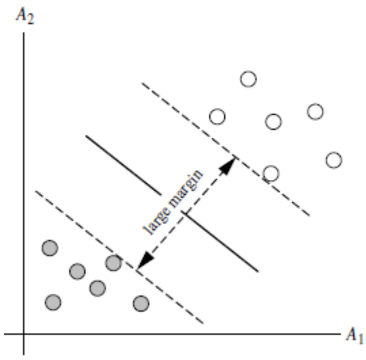
    3.1 例子：



两类？哪条线最好？

3.2 SVM寻找区分两类的超平面（hyper plane), 使边际(margin)最大



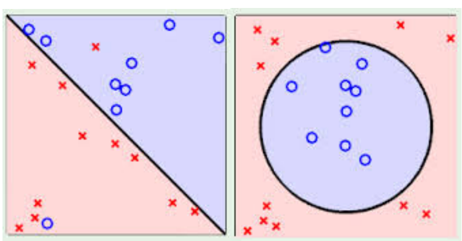


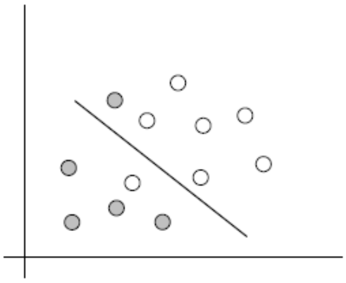
总共可以有多少个可能的超平面？无数条

如何选取使边际(margin)最大的超平面 (Max Margin Hyperplane)？

超平面到一侧最近点的距离等于到另一侧最近点的距离，两侧的两个超平面平行

3. 线性可区分(linear separable) 和 线性不可区分 （linear inseparable)

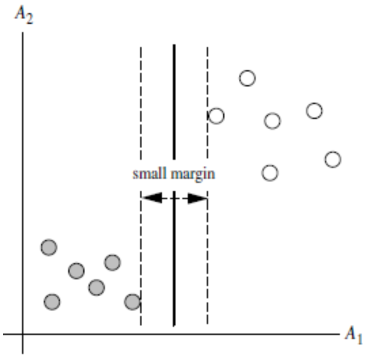


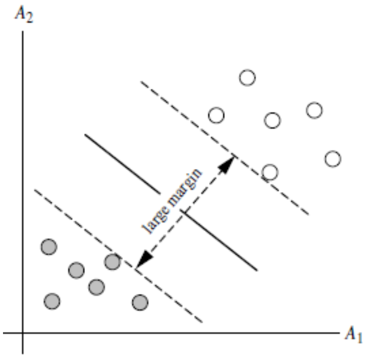


4. 定义与公式建立

超平面可以定义为：$W \cdot X + b = 0$

W: weight vectot, $W = \{w_1, w_2, \ldots, w_n\}$，n 是特征值的个数
X: 训练实例
b: bias





4.1 假设2维特征向量：X = (x1, X2)
把 b 想象为额外的 wight
超平面方程变为：$w_0 + w_1x_1 + w_2x_2 = 0.$

所有超平面右上方的点满足：$w_0 + w_1x_1 + w_2x_2 > 0.$
所有超平面左下方的点满足：$w_0 + w_1x_1 + w_2x_2 < 0$

调整weight，使超平面定义边际的两边：

$$H_1 : w_0 + w_1x_1 + w_2x_2 \geq 1 \text{ for } y_i = +1,$$
$$H_2 : w_0 + w_1x_1 + w_2x_2 \leq -1 \text{ for } y_i = -1$$

综合以上两式，得到：（1）
$y_i(w_0 + w_1x_1 + w_2x_2) \geq 1, \ \forall i.$

所有坐落在边际的两边的的超平面上的被称作"支持向量(support vectors)"

分界的超平面和H1或H2上任意一点的距离为 (i.e.: 其中||W||是向量的范数(norm))

If $W = \{w_1, w_2, \ldots, w_n\}$ then $\sqrt{W \cdot W} = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}$.

所以，最大边际距离为：

5. 求解

5.1 SVM如何找出最大边际的超平面呢(MMH)？

利用一些数学推倒，以上公式（1）可变为有限制的凸优化问题(convex quadratic optimization)

利用 Karush-Kuhn-Tucker (KKT)条件和拉格朗日公式，可以推出MMH可以被表示为以下"决定边界 (decision boundary)"

$$d(X^T) = \sum_{i=1}^{l} y_i \alpha_i X_i X^T + b_0,$$

其中，

是支持向量点 （ support vector)的类别标记 (class label)

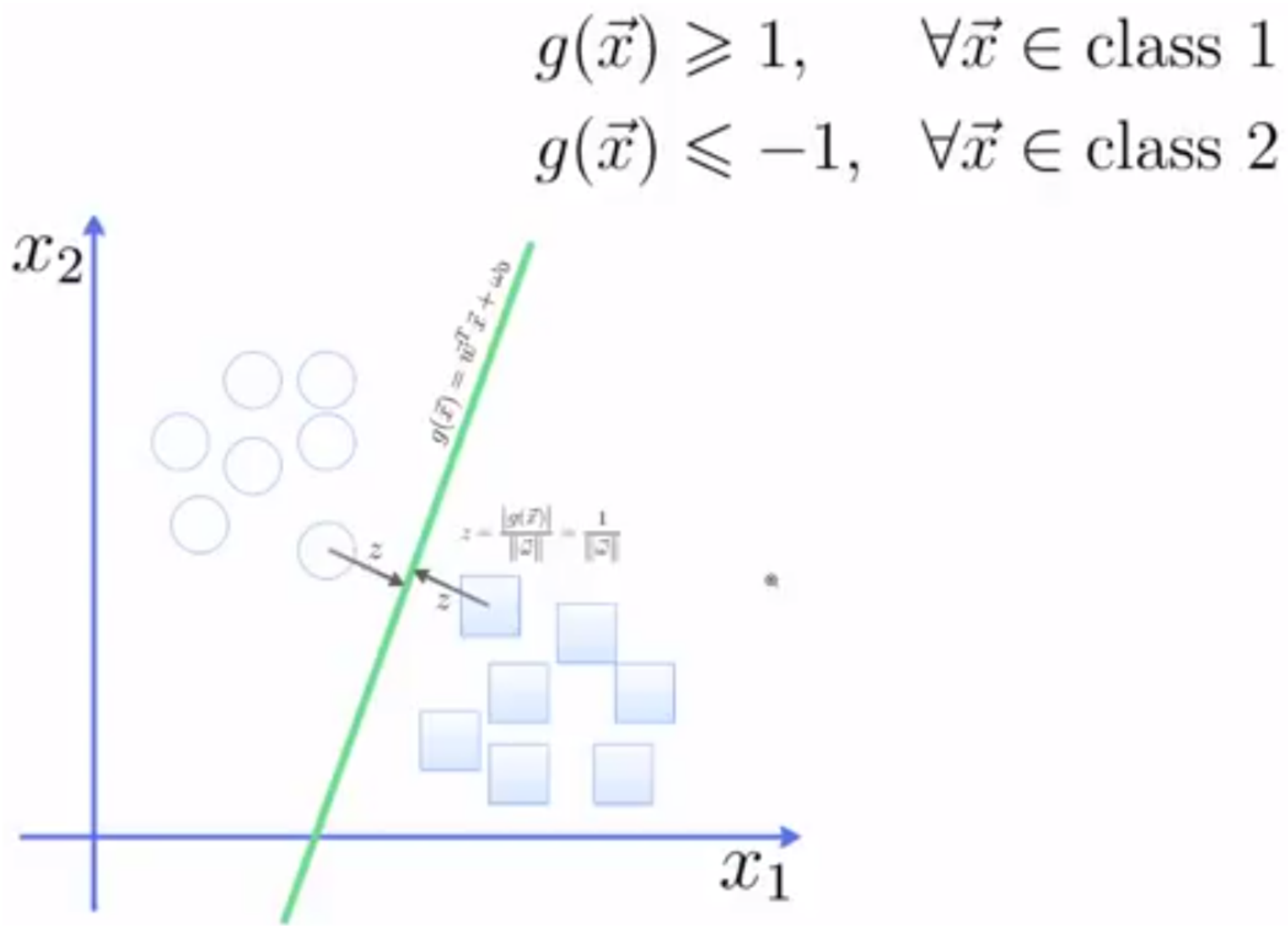
是要测试的实例

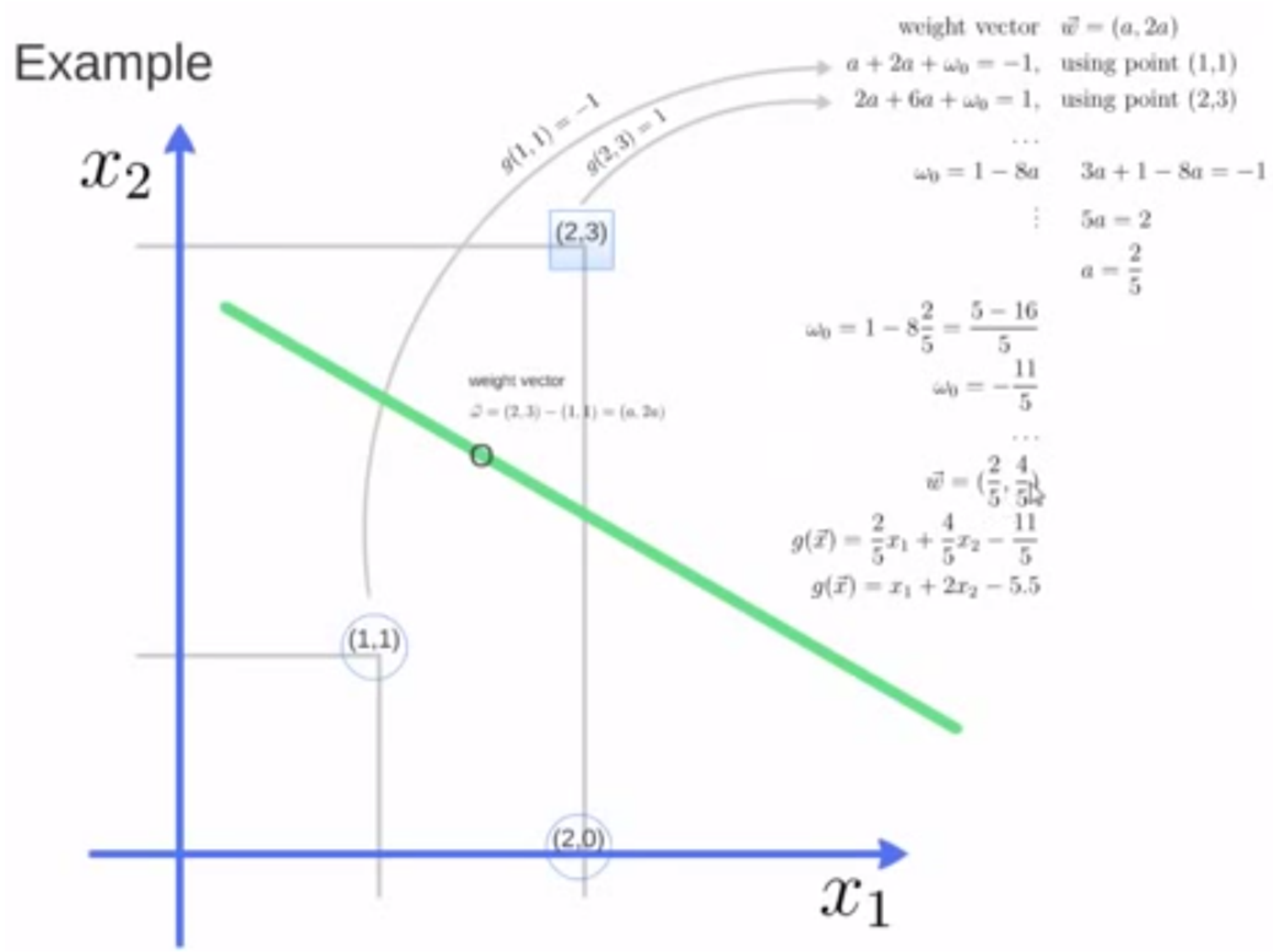
和 都是单一数值型参数，由以上提到的最有算法得出

是支持向量点的个数

5.2 对于任何测试 (要归类的) 实例，带入以上公式，得出的符号是正还是负决定

6. 例子：




Example
x_2
x_1
$g(1,1) = -1$
$g(2,3) = 1$
(2,3)
(1,1)
(2,0)
weight vector
$\vec{\omega} = (2,3) - (1,1) = (a, 2a)$
weight vector $\vec{w} = (a, 2a)$
$a + 2a + \omega_0 = -1$, using point (1,1)
$2a + 6a + \omega_0 = 1$, using point (2,3)
...
$\omega_0 = 1 - 8a$
$3a + 1 - 8a = -1$
$5a = 2$
$a = \frac{2}{5}$
$\omega_0 = 1 - 8\frac{2}{5} = \frac{5-16}{5}$
$\omega_0 = -\frac{11}{5}$
...
$\vec{w} = (\frac{2}{5}, \frac{4}{5})$
$g(\vec{x}) = \frac{2}{5}x_1 + \frac{4}{5}x_2 - \frac{11}{5}$
$g(\vec{x}) = x_1 + 2x_2 - 5.5$