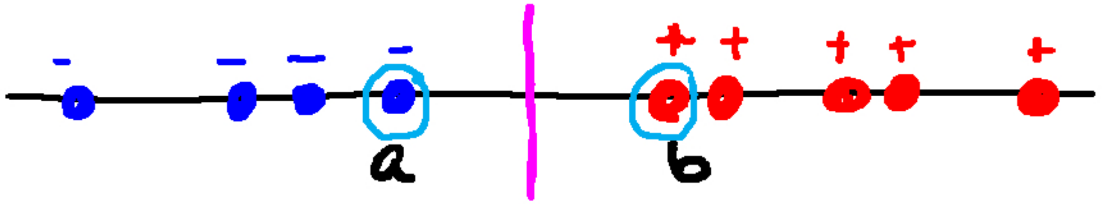
Support Vector Machines for Classification

Objectives

- Define margin for separable data
- Show support vector machines maximize margin
- Use hinge loss to define support vector machines for nonseparable data

Maximize margin for separable training data 2

Example:

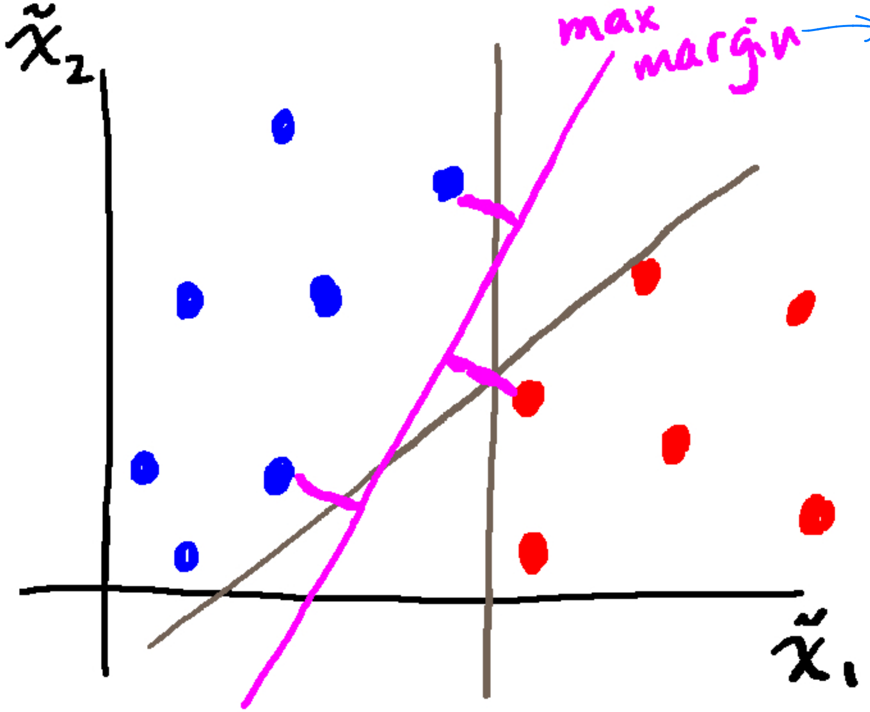


decision boundary?

margin: distance from boundary to nearest sample

max margin boundary: midpoint

only a, b matter



max margin (This is the line that separates the class equidistantly) decision boundary?

(☆ In order to shift the boundary away from the origin)

Data constant

feature $\underline{x}^T = [\,\underline{\tilde{x}}^T \quad 1\,]$

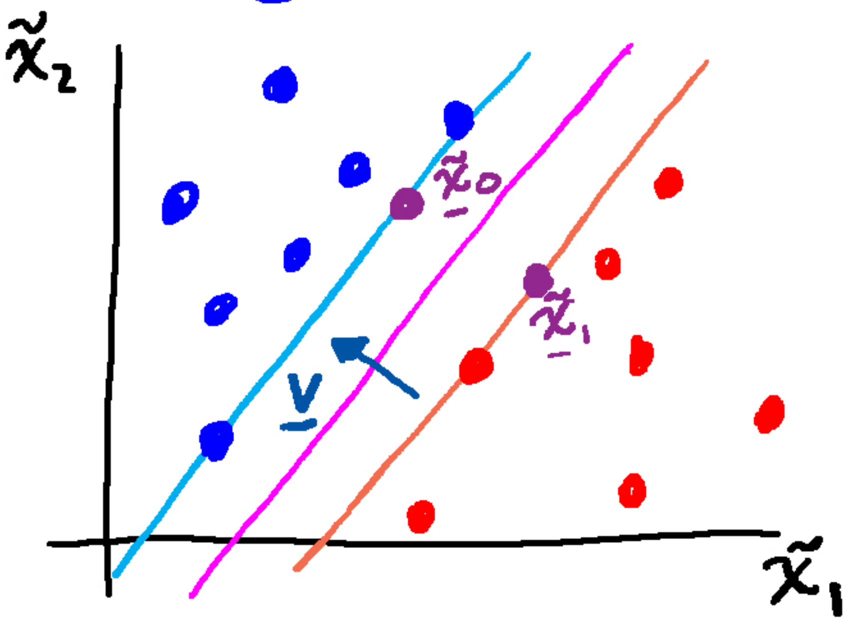
weights $\underline{w}^T = [\,\underline{\tilde{w}}^T \quad w_0\,]$

Constant associated with the offset above

decision $\hat{d} = \text{sign}(\underline{x}^T \underline{w})$

$$\hat{d} = \begin{cases} 1 & \underline{\tilde{x}}^T \underline{\tilde{w}} + w_0 > 0 \\ -1 & \underline{\tilde{x}}^T \underline{\tilde{w}} + w_0 < 0 \end{cases}$$

Margin is determined by $\|\tilde{\underline{w}}\|_2^{-1}$



label "-1": $\tilde{\underline{x}}^T\tilde{\underline{w}} + w_0 \le -1$

label "+1": $\tilde{\underline{x}}^T\tilde{\underline{w}} + w_0 \ge 1$

boundary : $\tilde{\underline{x}}^T\tilde{\underline{w}} + w_0 = 0$ In the middle

margin: ½ distance between / /
measure in direction $\underline{v}$

Unit normal to boundary plane: $\underline{v} = \dfrac{\tilde{\underline{w}}}{\|\tilde{\underline{w}}\|_2}$ (Normalized)

2m gives the magnitude, v gives the direction

Margin $m = \frac{1}{2}\|\tilde{\underline{x}}_1 - \tilde{\underline{x}}_0\|_2$ $\qquad \tilde{\underline{x}}_1 = \tilde{\underline{x}}_0 + 2m\,\underline{v}$

x_0^T is on the boundary of $x=-1$

but $\tilde{\underline{x}}_0^T\tilde{\underline{w}} + w_0 = -1$

so $2 = 2m\|\tilde{\underline{w}}\|_2$

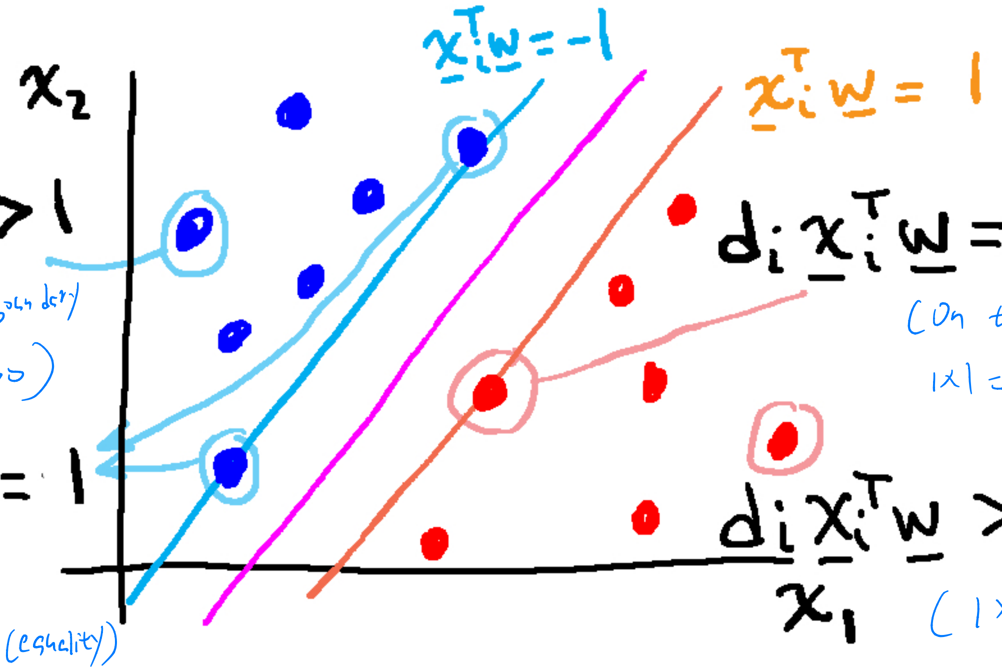
$$1 = \tilde{\underline{x}}_1^T\tilde{\underline{w}} + w_0 = \tilde{\underline{x}}_0^T\tilde{\underline{w}} + 2m\frac{\tilde{\underline{w}}^T}{\|\tilde{\underline{w}}\|_2}\tilde{\underline{w}} + w_0$$

x_1^T is on the boundary

$$\boxed{m = \|\tilde{\underline{w}}\|_2^{-1}}$$ Margin

Support Vector Machine maximizes margin

Correct classification: $d_i(\tilde{\underline{x}}_i^T \tilde{\underline{w}} + w_0) \geq 1$



$d_i \underline{x}_i^T \underline{w} > 1$

(Left of the boundary

$(-1) \times$ neg > 0)

$d_i \underline{x}_i^T \underline{w} = 1$

On the boundary (equality)

$(-1 \times -1 = 0)$

(On the boundary

$1 \times 1 = 1$)

($1 \times$ positive > 1)

SVM:

[max the margin ($\|\tilde{w}\|_2^{-1}$) is to min the $\|\tilde{w}\|_2^2$]

$$\underbrace{\min_{\tilde{\underline{w}}, w_0} \|\tilde{\underline{w}}\|_2^2}_{\text{max margin}} \quad \text{s.t.} \quad \underbrace{d_i(\tilde{\underline{x}}_i^T \tilde{\underline{w}} + w_0) \geq 1 \quad i = 1, 2, \ldots, N}_{\text{perfect classification}}$$

subject to

a set of liner constraints

(Every sample should be correct)

unique solution

Boundary defined by $\underline{x}_i$ for which $d_i \underline{x}_i^T \underline{w} = 1$

Called Support Vectors → 3 SVs here in the example

SVM for non separable data uses hinge loss

$$\min_{\underline{w}} \underbrace{\sum_{i=1}^{N} (1 - d_i \underline{x}_i^T \underline{w})_+}_{\text{loss}} + \lambda \underbrace{\|\underline{\tilde{w}}\|_2^2}_{\sim (\text{margin})^{-2}}$$

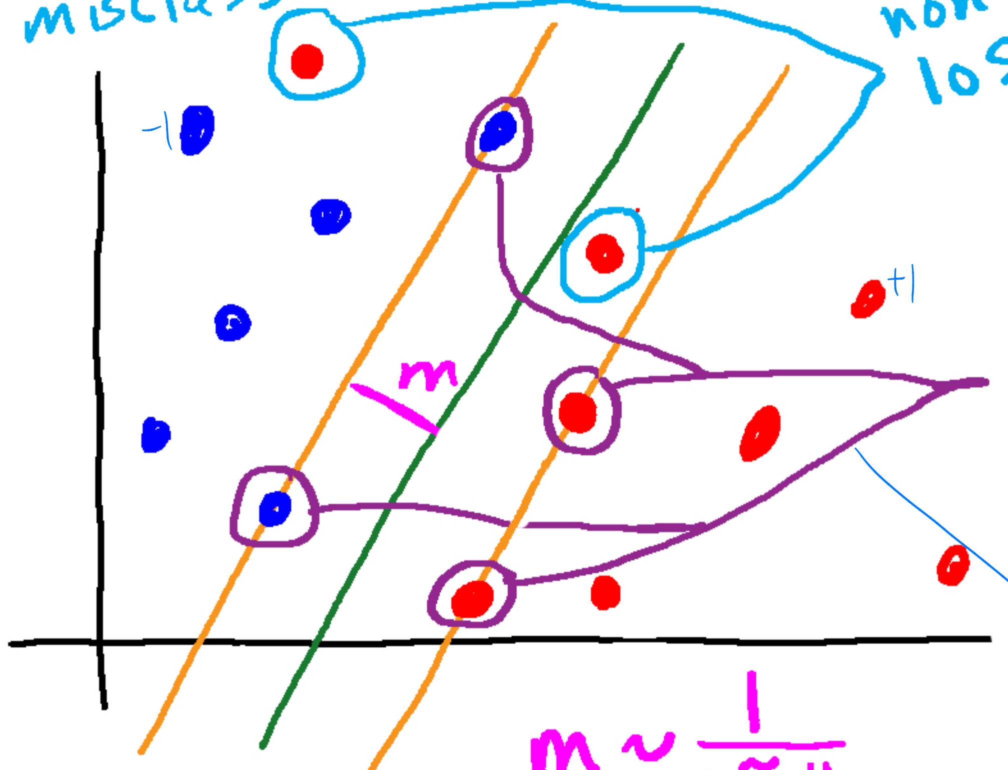
(ℓ_2 regularization)

最小化 $\|w\|^2$ 就是最大化 margin ,

misclassification loss

non zero loss

Contribute to the hinge loss



support vectors

这些点是 support vectors

软间隔 SVM,在最大化间隔的同时,也允许一些点违反间隔约束

$$m \sim \frac{1}{\|\underline{\tilde{w}}\|_2}$$

small λ → 用来控制正则化的强度

large λ