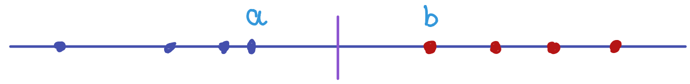
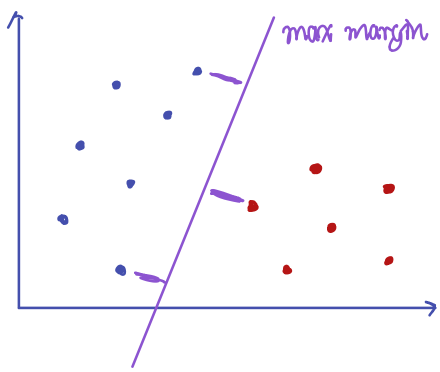
Maximize Margin for separable training data

Example:



Margin: distance from boundary to nearest sample

max margin boundary: midpoint, only a, b matter



decision boundary?

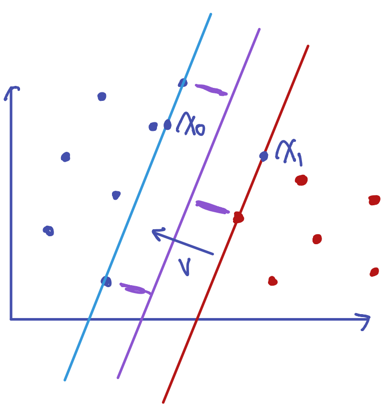
feature: $x^T = [\tilde{x}^T \; 1]$

weights: $w^T = [\tilde{w}^T \; w_0]$

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

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

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



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

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

boundary : $\tilde{x}^T\tilde{w} + w_0 = 0$

margin: $\frac{1}{2}$ distance between blue and red line measure in direction v

Unit norm to boundary plane: $v = \frac{\tilde{w}}{\|\tilde{w}\|_2}$

Margin $m = \frac{1}{2}\|\tilde{x}_1 - \tilde{x}_0\|_2$, $\tilde{x}_1 = \tilde{x}_0 + \underset{\text{distance}}{2m}\underset{\text{direction}}{v}$

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

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

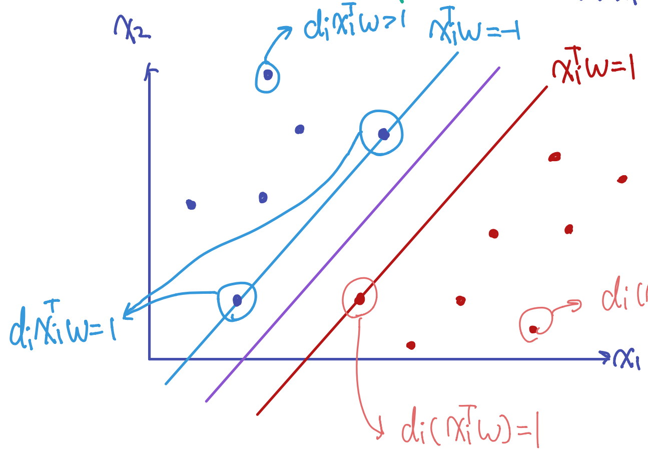
so $2 = 2m\frac{\tilde{w}^T}{\|\tilde{w}\|_2}\tilde{w}$

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

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

Support Vector Machine maximizes margin

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



SVM:

$$\underbrace{\min_{\tilde{w}, w_0} \|\tilde{w}\|_2^2}_{\text{max margin}} \text{ such that } \underbrace{d_i(\hat{x}_i^T\tilde{w}+w_0)\ge 1 \quad i=1,2,\cdots,N}_{\text{perfect classification}}$$

unique solution

Boundary defined by x_i for which $d_i x_i^T w = 1$, called Support Vectors

SVM for nonseparable data uses hinge loss

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

ℓ_2 regularization