EECS 127

Support Vector Machines 2

Admin

- Last lecture.
- Guest lecture Prof. El Ghaoui on Tuesday.
- Thursday, April 20 in class!
- Evaluations at end of class

Last time:

Hard-margin SVM

$$\min \frac{1}{2}\|\vec{w}\|_2^2$$

$$\text{s.t. } y_i(\vec{w}^T\vec{x}_i - b) \geq 1 \quad \forall i$$

Soft-margin SVM.

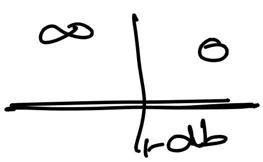
$$\min \frac{1}{2}\|\vec{w}\|_2^2 + C\cdot\sum_{i=1}^{n}\xi_i$$

$$\text{s.t. } y_i(\vec{w}^T\vec{x}_i - b) \geq 1 - \xi_i \quad \forall i$$

$$\xi_i \geq 0 \quad \forall i$$

$L_{0-\infty}$ losss

$$L_{0,\infty}(a,b) = \begin{cases} 0 & \text{if } 1-ab<0 \\ \infty & \text{if } 1-ab>0 \end{cases}$$



$$\min_{\vec{w},\vec{b}} \frac{1}{n}\sum L_{0,\infty}(y_i, \vec{w}^T\vec{x}_i - b) + \lambda\|\vec{w}\|_2^2$$

Hinge-loss formulation

$$\min_{\vec{w},b} \frac{1}{n}\sum_{i=1}^{n} L_{hinge}(y_i, \vec{w}^T\vec{x}_i - b) + \lambda\|\vec{w}\|_2^2$$

$$L_{hinge}(a,b) = \max(1-ab, 0)$$

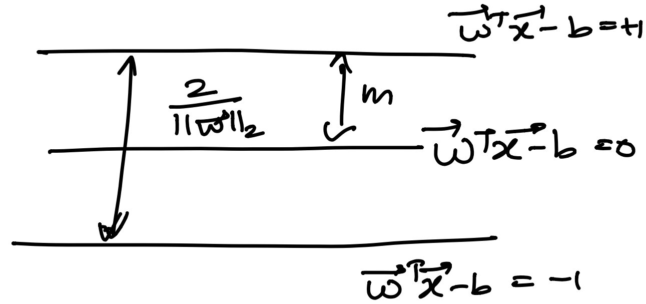
$$L_{hinge}(y_i, \vec{w}^T\vec{x}_i - b) = \max(1 - y_i(\vec{w}^T\vec{x}_i - b), 0)$$

Logistic Regression

$$\min_{\vec{w},\vec{b}} \frac{1}{n}\sum \underbrace{\ln\left(1 - e^{-(y_i(\vec{w}^T\vec{x}_i - b))}\right)}_{\text{logistic loss}}$$

$$y_i(\vec{w}\vec{x}_i - b) > 0 \Rightarrow \text{point is correctly classified}$$

max margin. $\boxed{m}$



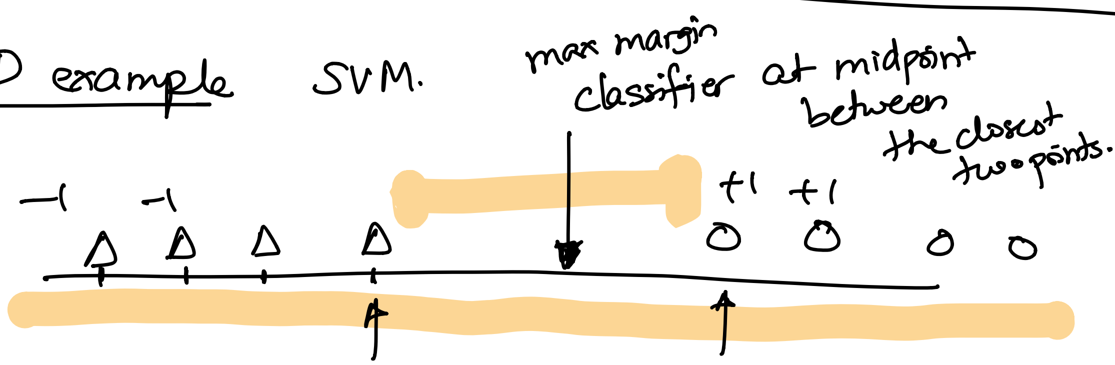
$$m = \frac{1}{\|\vec{w}\|_2}$$

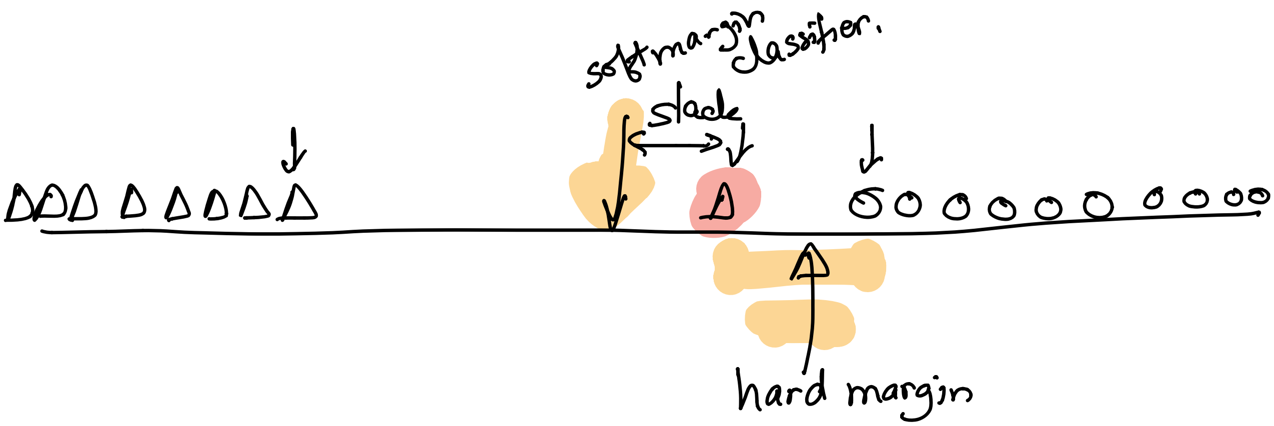
Constraint:

$$m\left(y_i\left(\vec{w}^T\vec{x}_i - b\right)\right) \geq (1 - \xi_i)\,m$$

$$\hookrightarrow \frac{y_i\left(\vec{w}^T\vec{x}_i - b\right)}{\|\vec{w}\|_2} \geq m - \underbrace{m \cdot \xi_i}$$

1D example SVM.





→ SPARSITY of test points that matter

→ Feature of SVMs.

<u>Dual perspective.</u> $\vec{\alpha}, \vec{\beta} \geq 0.$

$$L(\vec{\omega}, b, \vec{\xi}, \underbrace{\vec{\alpha}, \vec{\beta}}_{\text{dual}}) = \frac{1}{2}\|\vec{w}\|_2^2 + C \cdot \sum_{i=1}^{n} \xi_i$$

$$+ \sum_{i=1}^{n} \alpha_i \left((1-\xi_i) - y_i(\vec{\omega}^T\vec{x}_i - b)\right)$$

$$+ \sum_{i=1}^{n} \beta_i(-\xi_i)$$

$$= \frac{1}{2}\|\vec{w}\|_2^2 - \sum_{i=1}^{n} \alpha_i y_i (\vec{\omega}^T\vec{x}_i - b) + \sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} (C - \alpha_i - \beta_i)\xi_i$$

<u>Recall:</u>

$$p^* = \min_{\vec{x}} \max_{\vec{\lambda}} L(\vec{x}, \vec{\lambda})$$

$$d^* = \max_{\vec{\lambda}} \min_{\vec{x}} L(\vec{x}, \vec{\lambda})$$

When strong duality holds we can interchange the min and the max, because $p^* = d^*$

Our problem: · Convex · Affine $\Big\} \Rightarrow$ Strong duality holds

$\Rightarrow$ KKT conditions are necessary + sufficient.

---

① First-order conditions

$$\nabla_{\vec{w}} L(\quad) = \vec{w} - \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i = 0$$

$$\boxed{\vec{w}^* = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i}$$

opt. $\vec{w}$ can be expressed in terms of opt. dual vars and training data.

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} \alpha_i y_i = 0$$

(Weighted sum of + points and − points should be 0)

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \beta_i = 0$$

Substitute back into Lagrangian.

$$\cancel{\mathcal{L}}\, L(\vec{w}, b, \vec{\xi}, \vec{\alpha}, \vec{\beta})$$

$$\cancel{\sum \underbrace{(C - \alpha_i - \beta_i)}_{0} \xi_i}$$

$$= \frac{1}{2}\|\vec{w}\|_2^2 - \sum \alpha_i y_i (\vec{w}^T \vec{x_i} - b) + \sum \alpha_i + 0$$

$$= \frac{1}{2}\|\vec{w}\|_2^2 - \sum_{i=1}^{n} \alpha_i y_i \vec{w}^T \vec{x_i} + 0 + \sum_{i=1}^{n} \alpha_i$$

$$(b \cdot \underbrace{\sum \alpha_i y_i}_{0})$$

$$= \frac{1}{2}\vec{W}^T \vec{W} - \sum_{i=1}^{n} \alpha_i y_i \vec{x_i}^T \vec{w} + \sum_{i=1}^{n} \alpha_i$$

$$= \left(\frac{1}{2}\vec{W}^T - \sum_{i=1}^{n} \alpha_i y_i \vec{x_i}^T\right)\vec{w} + \sum \alpha_i$$

$$\vec{w}^* = \sum_{i=1}^{n} \alpha_i y_i \vec{x_i}$$

$$= -\frac{1}{2}\vec{W}^T \cdot \vec{W} + \sum_{i=1}^{n} \alpha_i$$

$$= -\frac{1}{2}\left(\sum_{i=1}^{n} \alpha_i y_i \vec{x_i}^T\right)\left(\sum \alpha_i y_i \vec{x_i}\right) + \sum \alpha_i$$

↳ no more primal variables)

$$X = \begin{bmatrix} - \vec{x}_1^T - \\ - \vec{x}_2^T - \\ \vdots \\ - \vec{x}_n^T - \end{bmatrix}$$

$$= -\frac{1}{2}\,[\alpha_1 y_1 \;\; \alpha_2 y_2 \cdots \alpha_n y_n]\, X \cdot X^T \begin{bmatrix} \alpha_1 y_1 \\ \alpha_2 y_2 \\ \vdots \\ \alpha_n y_n \end{bmatrix} + \sum \vec{\alpha_i}$$

$$= -\frac{1}{2} \vec{\alpha}^T \underbrace{diag(y_i)\, X X^T diag(y_i)}_{Q} \vec{\alpha} + \sum \vec{\alpha}_i$$

$$= -\frac{1}{2} \vec{\alpha}^T Q \cdot \vec{\alpha} + \sum \vec{\alpha}_i$$

$$d^* = \max_{\vec{\alpha}, \vec{\beta}} \min_{\vec{w}, \vec{b}} \cdot L(\quad)$$

$$= \max_{\vec{\alpha} \geq 0, \vec{\beta} \geq 0} \vec{\alpha}^T Q \vec{\alpha} + \sum \alpha_i$$

$$\sum \alpha_i y_i = 0$$

$$\left.\begin{array}{l} C - \alpha_i - \beta_i = 0 \\ \alpha_i \geq 0 \\ \beta_i \geq 0 \end{array}\right] \Rightarrow \begin{array}{l} \beta_i = C - \alpha_i \geq 0 \\ 0 \leq \alpha_i \leq C \end{array}$$

$$\boxed{\begin{array}{l} = \max_{\vec{\alpha}} -\frac{1}{2} \vec{\alpha}^T Q \cdot \vec{\alpha} + \sum \alpha_i \qquad QP \\ \text{s.t } \sum_{i=1}^{m} \alpha_i y_i = 0 , \qquad 0 \leq \alpha_i \leq C. \end{array}}$$

Complementary slackness conditions

① $\alpha_i \left((1-\xi_i) - y_i (\vec{w}^T \vec{x}_i - b) \right) = 0$

② $\beta_i \xi_i = 0$

Reminder: $C - \alpha_i - \beta_i = 0$

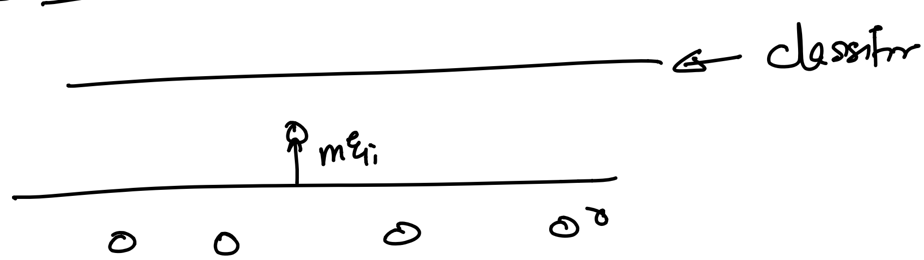
Consider

① If $\alpha_i = 0 \Rightarrow C - 0 - \beta_i = 0$

$\Rightarrow C = \beta_i$

$C \neq 0$ constant from regularizer

$\xi_i = 0$ because $\beta_i \neq 0$ from ②



ith point has no margin violation!

(2) $\alpha_i \neq 0$

(2a) $\alpha_i = C$ $\qquad\qquad C - \alpha_i - \beta_i = 0$

$\Rightarrow \beta_i = 0 \rightarrow$ Can't say anything about ξ_i

But

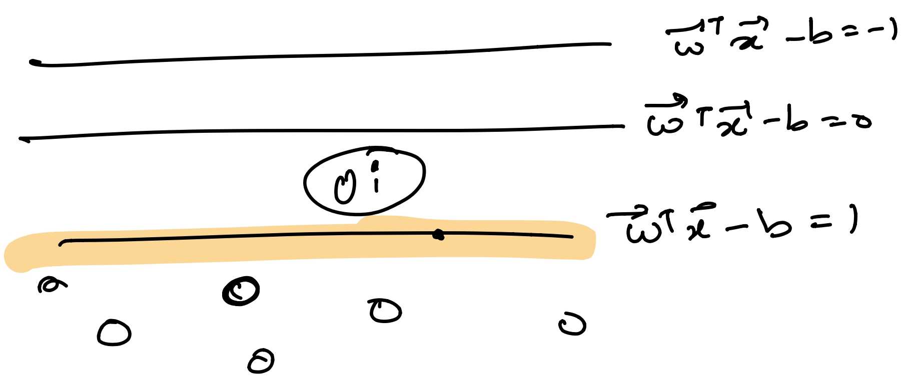
$$(1-\xi_i) - y_i(\vec{w}^T \vec{x}_i - b) = 0$$

$$\Rightarrow \quad y_i(\vec{w}^T \vec{x}_i - b) = 1 - \xi_i \leq 1$$

$$\xi_i \geq 0$$

Case $y_i = +1$ $\qquad$ point $+1$ want $\vec{w}^T \vec{x}_i - \vec{b} \geq 1$

$\Rightarrow (\vec{w}^T \vec{x}_i - b) \leq 1$ because $y_i = +1$



ith point is either on the margin or it ~~viola~~ violates the constraint.

③ $\alpha_i \neq 0, \quad 0 < \alpha_i < C$

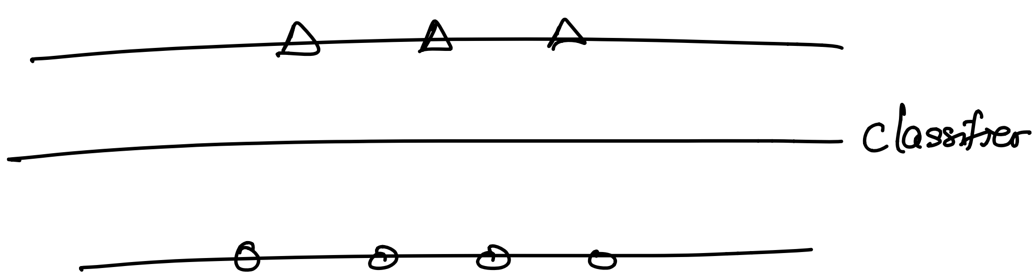
$$C - \alpha_i - \beta_i = 0$$

$$\beta_i \neq 0 \implies \xi_i = 0$$

$\implies$ No margin violation!

① $\implies y_i(\vec{w}^T \vec{x}_i - b) = 1 - \cancel{\xi_i}^{\,0}$

$$y_i(\vec{w}^T x_i - b) = 1$$



Support vectors.

WHY DID WE BOTHER?

$$\vec{w} = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i$$

Only the points that have non-zero dual variables matter! $\rightarrow$ Non-zero α_i points are called SUPPORT vectors

$$\boxed{\vec{w}^T \vec{x} - b}$$

How to find b?

Choose a point s.t. $0 < \alpha_i < C$

Use $y_i(\vec{w}^T \vec{x}_i - b) = 1$ to compute b

Dual:

$$\underbrace{Q} = \operatorname{diag}(Y) \underbrace{X X^T}_{\text{inner products}} (\operatorname{diag} Y)$$

$$\begin{bmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_n \end{bmatrix}$$

$$n \; \underbrace{[X X^T}_{} \; n$$

eg. $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$.

$$\phi_2(\vec{x}) = \begin{bmatrix} x_1^2 \\ \sqrt{2}\, x_1 x_2 \\ x_2^2 \end{bmatrix}$$

$$\phi_2(\vec{x})\,\phi_2(\vec{z}) = x_1^2 z_1^2 + 2x_1x_2z_1z_2 + x_2^2 z_2^2$$

$$= \underline{(\vec{x}^T\vec{z})^2}$$

"Kernel trick"