Lecture # 18

Discrimination / Classification / Separation Problem

Given 2 sets of data pt-s in $\mathbb{R}^n$

$\{\underline{x}_1, \dots, \underline{x}_N\}$ & $\{\underline{y}_1, \dots, y_M\}$

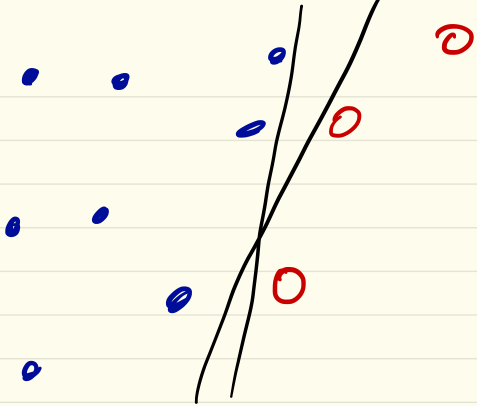
Find a f$\stackrel{n}{=}$ $f : \mathbb{R}^n \longmapsto \mathbb{R}$
(within certain class of f$\stackrel{n}{=}$s)
such that

$$f(x_i) > 0, \quad i = 1, \dots, N$$

$$f(y_i) < 0, \quad i = 1, \dots, M$$

The the f$\stackrel{n}{=}$ f (more carefully, the 0-level set of f, i.e. $\{\underline{x} \mid f(\underline{x}) = 0\}$)

discriminates/classifies/separates the two sets of data.



Simple case:

linear classifiers/
discriminator

$$f(\underline{x}) = \underline{a}^T \underline{x} - b$$

(i.e.) $\underline{a}^T \underline{x}_i - b > 0, \ \forall i = 1, \ldots, N$

& $\underline{a}^T \underline{y}_i - b < 0, \ \forall i = 1, \ldots, M$

$\Leftrightarrow$ Find a separating hyperplane.

Theorem of alternatives for linear inequalities

(text section 5.8.3)

The set of previous linear inequalities will NOT hold:

$\exists\ \underline{\lambda}, \underline{\tilde{\lambda}}$ s.t.

$$\underline{\lambda} \succcurlyeq 0,\ \underline{\tilde{\lambda}} \succcurlyeq 0,\ (\underline{\lambda}, \underline{\tilde{\lambda}}) \neq 0,\ \sum_{i=1}^{N} \lambda_i x_i = \sum_{i=1}^{M} \tilde{\lambda}_i y_i,$$

$$\boxed{\mathbb{1}^T \underline{\lambda} = \mathbb{1}^T \underline{\tilde{\lambda}}.}$$

$$\Updownarrow$$

normalize $\underline{\lambda}$ (i.e. set $\mathbb{1}^T \underline{\lambda} = 1$)

$$\Updownarrow$$

$$\underline{\lambda} \succcurlyeq 0,\ \underline{\tilde{\lambda}} \succcurlyeq 0,\ \mathbb{1}^T \underline{\lambda} = 1,\ \mathbb{1}^T \tilde{\lambda} = 1,\ \sum_{i=1}^{N} \lambda_i x_i = \sum_{i=1}^{M} \tilde{\lambda}_i y_i$$

$\Leftrightarrow$ (Geometric meaning) $\exists$ a pt. in $\mathrm{conv}\{\underline{x}_1, \dots, \underline{x}_N\}$ & $\mathrm{conv}\{\underline{y}_1, \dots, \underline{y}_M\}$

$\Leftrightarrow$ linearly separable iff convex hulls do NOT intersect.

Robust linear discriminator / Classifier

$\underbrace{\text{Robust linear discriminator}}_{\text{(RLD)}}$

Find optimal $(\underline{a}, b)$ in $f(\underline{x}) = \underline{a}^T \underline{x} - b$

(i.e.) one that maximizes the "gap" betn.

> 0 values @ x_i, & < 0 values @ y_i.

$$\Longleftrightarrow \left\{ \begin{array}{ll} \underset{t,\ \underline{a},\ b}{\text{maximize}} & t \\ \text{s.t.} & \underline{a}^T \underline{x}_i - b \geq t, \quad i = 1, \ldots, N \\ & \underline{a}^T \underline{y}_i - b \leq -t, \quad i = 1, \ldots, M \\ & \|\underline{a}\|_2 \leq 1 \end{array} \right.$$

$$\text{argmax} \rightarrow (t^*, \underline{a}^*, b^*)$$

linearly separable $\Longleftrightarrow$ $t^* > 0$

Can prove: $\|\underline{a}^*\|_2 = 1$

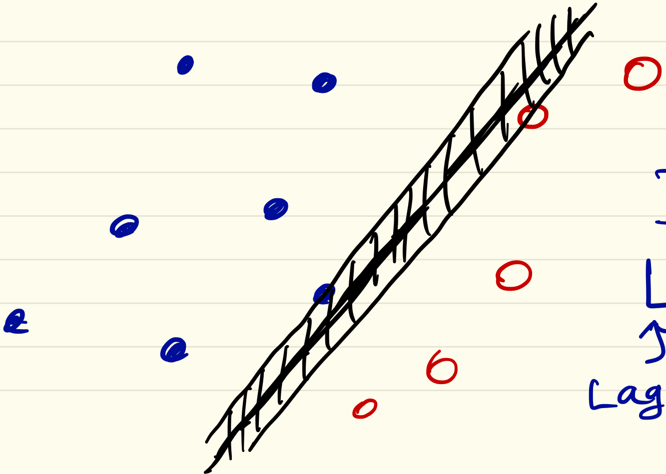
If $\|\underline{a}\|_2 = 1$, then $(\underline{a}^T \underline{x}_i - b)$ is the Euclidean dist. from $\underline{x}_i$ to the sep. hyp. plane $\mathcal{H} = \{\underline{z} \in \mathbb{R}^n \mid \underline{a}^T \underline{z} = b\}$

Similarly $(b - a^T \underline{y}_i)$ is dist. from y_i to $\mathcal{H}$.

$\therefore$ The RLD finds the maximal separator

$\Updownarrow$

Thickest "slab" between 2 datasets



Dual of the RLD primal

$$L = -t + \sum_{i=1}^{N} u_i \left(t + b - \underline{a}^T \underline{x}_i\right) + \sum_{i=1}^{M} v_i \left(t - b + \underline{a}^T \underline{y}_i\right) + \lambda\left(\|\underline{a}\|_2 - 1\right)$$

Lagrangian

The Lagrangian $L(\underbrace{\underline{a}, b, t}_{\text{Primal variables}}, \underbrace{\underline{u}, \underline{v}, t}_{\text{Dual variables}})$

Dual function: $g(\underline{u}, \underline{v}, \lambda) = \inf_{\underline{a}, b, t} L(\underline{a}, b, t, \underline{u}, \underline{v}, t)$

$$= \inf_{\underline{a}} \inf_{b, t} L(\underline{a}, b, t, \underline{u}, \underline{v}, t)$$

$$= \inf_{\underline{a}} \left\{ \underline{a}^T \left[\sum_{i=1}^{M} v_i y_i - \sum_{i=1}^{N} u_i x_i \right] + \lambda \| \underline{a} \|_2 - \lambda \right\}$$

$$= \begin{cases} -\lambda & \text{if } \left\| \sum_{i=1}^{M} v_i y_i - \sum_{i=1}^{N} u_i x_i \right\|_2 \leq \lambda \\ -\infty & \text{otherwise} \end{cases}$$

Conditions

$$\begin{cases} \mathbb{1}^T \underline{u} = 1/2 \\ \mathbb{1}^T \underline{v} = 1/2 \end{cases}$$

$\therefore$ The dual problem :

$$\underset{}{\text{maximize}} \quad - \left\| \sum_{i=1}^{M} v_i y_i - \sum_{i=1}^{N} u_i x_i \right\|_2$$

$$\text{s.t.} \qquad \underline{u} \succcurlyeq \underline{0}, \quad \mathbb{1}^T \underline{u} = 1/2,$$

$$\underline{v} \succcurlyeq \underline{0}, \quad \mathbb{1}^T \underline{v} = 1/2 .$$

Think of $2 \sum_{i=1}^{N} u_i x_i$ as a point in $\text{conv}\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N\}$

and $2 \sum_{i=1}^{M} v_i y_i$ as " " $\text{conv}\{\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_M\}$

$\therefore$ Dual objective = minimize (half) the distance between these 2 points

$\Updownarrow$

Find (half) the distance between $\text{conv}\{\underline{x}_1, \ldots, \underline{x}_N\}$ and $\text{conv}\{\underline{y}_1, \ldots, \underline{y}_M\}$

SVM (Support Vector Machine) / Support Vector Classifier

What if the dataset cannot be linearly separable?

Approximate linear classifier

$\Updownarrow$

minimize # of misclassification error

(difficult combinatorial optimization)

<u>Recall</u>: linear classifier:

$$a^T x_i - b > 0, \quad i = 1, \dots, N$$

$$a^T y_i - b < 0, \quad i = 1, \dots, M$$

$$\Longleftrightarrow \quad \underline{a}^T x_i - b \geq 1, \; i = 1, \dots, N, \quad \underline{a}^T y_i - b \leq -1, \; i = 1, \dots, M$$

relax for approximate classification

Introduce:

$$\left.\begin{array}{l} u_1, \dots, u_N \geq 0 \\ \& \ v_1, \dots, v_M \geq 0 \end{array}\right\} \text{s.t.} \quad \begin{array}{l} \underline{a}^T \underline{x}_i - b \geq 1 - u_i, \quad i=1,\dots,N \\ \underline{a}^T \underline{y}_i - b \leq -(1 - v_i), \quad i=1,\dots,M \end{array}$$

$\underline{u}, \underline{v} \equiv 0 \iff$ original LD problem.

$\therefore$ $\underline{u}, \underline{v}$ measure how much the inequalities are violated.

Heuristic

$$\underset{\underline{u}, \underline{v}, \underline{a}, b}{\text{minimize}} \quad \underbrace{\mathbb{1}^T \underset{N\times 1}{\underline{u}} + \mathbb{1}^T \underset{M\times 1}{\underline{v}}}_{\text{total \# of violation}}$$

relaxation

$$\text{s.t.} \quad \underline{a}^T \underline{x}_i - b \geq 1 - u_i, \quad \underline{a}^T \underline{y}_i - b \leq -(1 - v_i), \quad \underline{u} \geq 0, \underline{v} \geq 0$$

We can mix RLD & SVM:

Recall, width of the $\underbrace{\text{slab}}$ $= 2 / \|\underline{a}\|_2$

$\rightarrow \{\underline{z} \in \mathbb{R}^2 \mid -1 \leq \underline{a}^T \underline{z} - b \leq +1\}$

$$\therefore \min_{\underline{a}, b, \underline{u}, \underline{v}} \|\underline{a}\|_2 + \gamma \underbrace{(\mathbb{1}^T u + \mathbb{1}^T v)}_{\text{misclassification error}}$$

regularization

$$\text{s.t.} \quad a^T \underline{x}_i - b \geq 1 - u_i, \quad i = 1, \dots, N$$

$$a^T \underline{y}_i - b \leq -(1 - v_i), \quad i = 1, \dots, M$$

$$\underline{u} \succeq 0, \quad \underline{v} \succeq 0,$$

Nonlinear discrimination :

e.g. polynomial s.t. $f(x_i) > 0$, $i = 1, \dots, N$

$f(y_i) < 0$, $i = 1, \dots, M$

If f is a polynomial on $\mathbb{R}^n$ with degree $\leq d$

(i.e.) $$f(\underline{x}) = \sum_{i_1 + \dots + i_n \leq d} a_{i_1 \dots i_d} \, x_1^{i_1} \cdots x_n^{i_n}$$

Geometric meaning: whether 2 sets can be separated by an algebraic surface

cvx examples