Lecture 3

Today: mostly examples

Last time: ℓ_1, ℓ_∞ norm minimization – equivalent to LP

$x \in \mathbb{R}^n$, $\quad \min_x \|x\|_\infty \quad$ ℓ_∞ infinity norm

$$\|x\|_\infty = \max\{|x_1|, |x_2|, \dots, |x_n|\}$$

$$= \max\{x_1, -x_1, x_2, -x_2, \dots, x_n, -x_n\}$$

$$\begin{array}{lll} \min & t & \\ \text{st} & x_i \le t & i = 1, \dots, n \\ & -x_i \le t & \end{array}$$

ℓ_1-norm $\quad \min_x \|x\|_1$

$x \in \mathbb{R}^n$

$$\|x\|_1 = \sum_{i=1}^n |x_i| = \sum_{i=1}^n \max\{x_i, -x_i\}$$

$$\min_{x, t_1, \dots, t_n} t_1 + t_2 + \dots + t_n$$

$$\text{s t} \quad \left.\begin{array}{r} x_i \le t_i \\ -x_i \le t_i \end{array}\right\} \; -t_i \le x_i \le t_i \quad i = 1, \dots, n.$$

$$\min \; 1^T t \qquad t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}$$

$$\text{s t} \quad -t \le x \le t$$

- other norms: use approximation

$$A \|x\|_p \le \|x\|_q \le B \|x\|_p$$

<u>Example</u> $\quad \|x\|_2 \le \|x\|_1 \le \sqrt{n}\, \|x\|_2$

1) $\|x\|_2^2 = \sum_{i=1}^n x_i^2 = \sum_{i=1}^n |x_i|^2 \le \sum |x_i|^2 + 2 \sum_{i \ne j} |x_i| |x_j|$

$$= \left(\sum |x_i| \right)^2 = \| x_i \|_1^2$$

2) $\|x\|_1 = \sum_{i=1}^n |x_i| \le \sqrt{n} \left(\sum |x_i|^2 \right)^{1/2} = \sqrt{n}\, \|x\|_2$

$\uparrow$ Cauchy-Swartz

P1	P2
$\min \|x\|_2$ s t some constraints $Ax \leq b$	$\min \|x\|_1$ s t same constraints $Ax \leq b$
Optimal value p^* at x_0^*	Optimal value q^* at x_1^*

How far is q^* from p^* ?

Use $\|x\|_2 \leq \|x\|_1$ + any feasible solution gives an upper bound to the min value

$\|x\|_1 \leq \sqrt{n}\,\|x\|_2$

x_0^* is feasible in P2 $\Rightarrow q^* \leq \|x_0^*\|_1 \leq \sqrt{n}\,\|x_0^*\|_2$

$= \sqrt{n}\,p^*$

x_1^* is feasible in P_1 $\Rightarrow p^* \leq \|x_1^*\|_2 \leq \|x\|_1 = q^2$

$$\underbrace{p^* \leq q^* \leq \sqrt{n}\,p^*}_{\sqrt{n} \text{ approximation algorithm}}$$

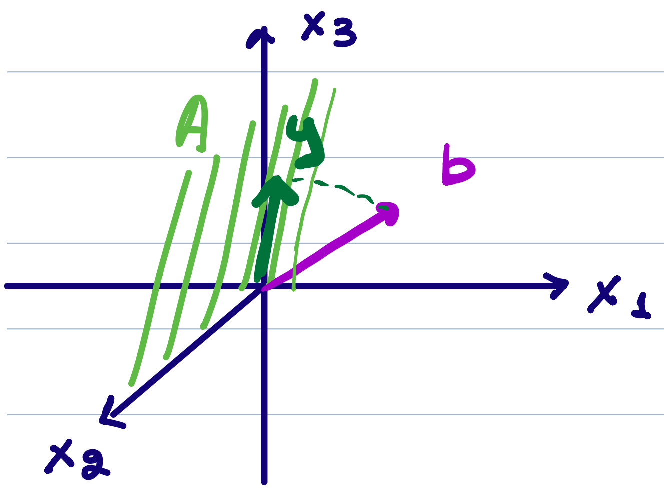
"Difference" between $\| \|_1$ and $\| \|_2$

- small entries contribute more to $\| \|_1$ and less to $\| \|_2$
  $x = (0.1 \ , \ 10)$

- large entries contribute more to $\| \|_2$

Error measurement
$\ell_2 \rightarrow$ lots of small errors
$\ell_1 \rightarrow$ most entries zero and a few larger ones

Example 1 : Assume we are given a vector $b \in \mathbb{R}^n$. We want to find the "closest" in the $\| \|_\infty$ sense, vector y, inside a subspace spanned by the columns of a matrix A.



$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_k \\ | & | & & | \end{bmatrix}$$

$$y = \begin{matrix} | \\ a_1 \\ | \end{matrix} x_1 + \begin{matrix} | \\ a_2 \\ | \end{matrix} x_2 + \cdots + \begin{matrix} | \\ a_n \\ | \end{matrix} x_n$$

$$y = A \cdot x$$

$n \times 1 \quad n \times k \quad k \times 1$

variables: x, y

obj. function: $\|b - y\|_\infty$

constraints: $y = Ax$

$$\begin{aligned} \min \ & \|y - b\|_\infty \\ \text{st} \ & y = Ax \end{aligned} \iff \min \|Ax - b\|_\infty$$

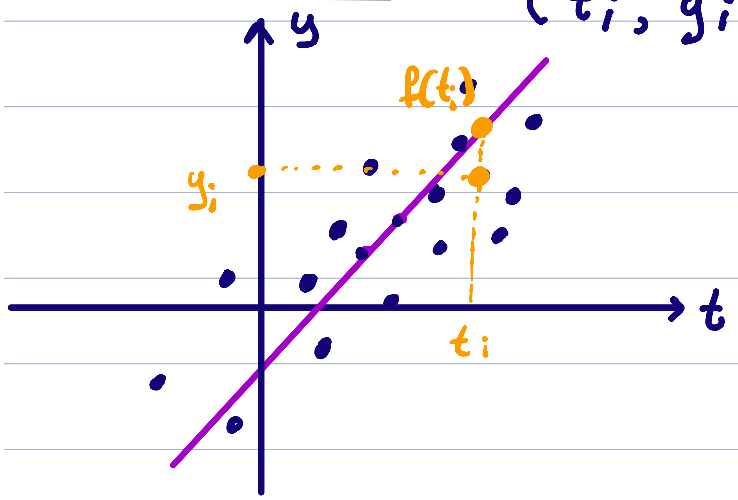
$$\iff \begin{aligned} \min \ & \|z\|_\infty \\ & z = Ax - b \end{aligned}$$

$$\iff \begin{aligned} \min \ & t \\ \text{st} \ & -t \le z_i \le t \\ & z = Ax - b \end{aligned}$$

$$\iff \begin{aligned} & \min \ t \\ & -t \cdot \mathbb{1} \le Ax - b \le t \cdot \mathbb{1} \end{aligned}$$

Example 2

We are given m points $(t_i, y_i) \in \mathbb{R}^2$



We want to find an affine function $f(t) = at + b$ such that, for every t_i, $f(t_i)$ is close to y_i in ℓ_1 distance

Unknowns: a, b

Obj. function

$$\min \left\| \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} - \begin{pmatrix} f(t_1) \\ \vdots \\ f(t_m) \end{pmatrix} \right\|_1$$

$$\min_{a,b} \left\| \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \\ 1 & t_m \end{pmatrix} \begin{pmatrix} b \\ a \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \right\|_1$$

min of the form

$$\min_x \|Ax - b\|_1$$

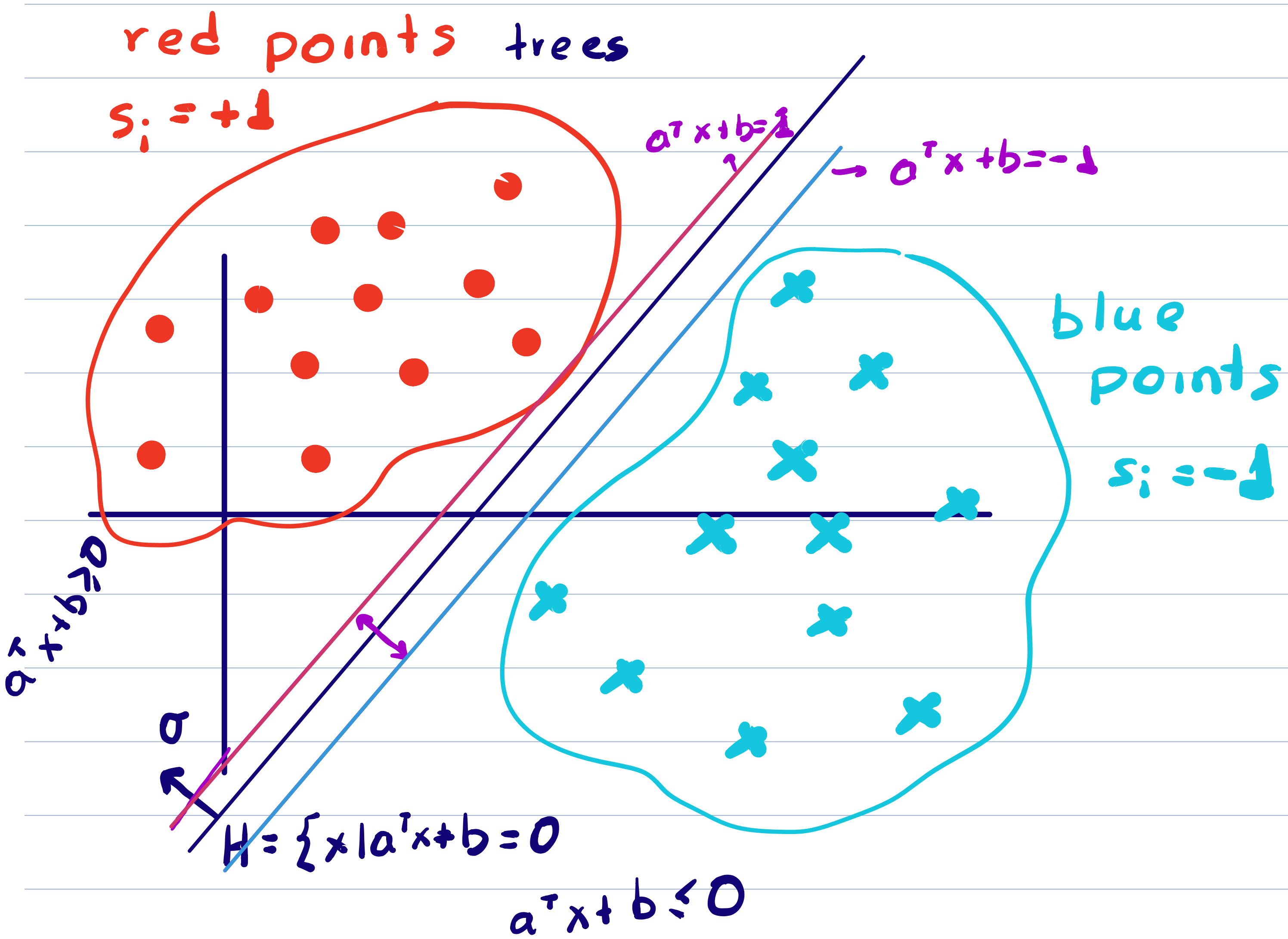
$$\min \; z_1 + z_2 + \cdots + z_m$$
$$\text{st} \quad |b + a t_i - y_i| \leq z_i \quad i = 1, \dots, m$$

Example 3 Linear Classification

Consider a set of points $x_1, x_2, \dots, x_m \in \mathbb{R}^n$

Each point comes with an associated label $s_i \in \{+1, -1\}$

training data (x_i, s_i)



Find a hyperplane $H = \{x \mid a^T x + b = 0\}$

that separates "as well as possible" the trainining points

We require that:

$s_i = -1 \qquad a^T x_i + b \leq -1$ (*)

$s_i = +1 \qquad a^T x_i + b \geq +1$

- We use +1 and −1 to avoid the trivial solution $a = 0$, $b = 0$
- I could have used any other constant instead of ± 1

Unknowns: a, b

"hinge loss"

Obj. function: $\sum_i \max\{0, 1 - s_i(a^T x_i + b)\}$

Captures a "penalty" of violating (*)

- If $s_i = +1$ $\qquad a^T x_i + b \geq 1$ → if this holds 0 penalty

→ if it doesn't hold then

$a^T x_i + b < 1$

penalty $1 - (a^T x_i + b)$

- If $s_i = -1$, $\qquad a^T x_i + b \leq -1$ → if this holds we get 0 penalty

→ if not

$a^T x_i + b > -1$ $\qquad$ penalty:

$a^T x_i + b + 1$

joint penalty $\quad 1 - s_i(a^T x_i + b)$

$$\min_{a,\, b} \sum_{i=1}^{n} \max\{0,\ 1 - s_i(a^T x_i + b)\}$$

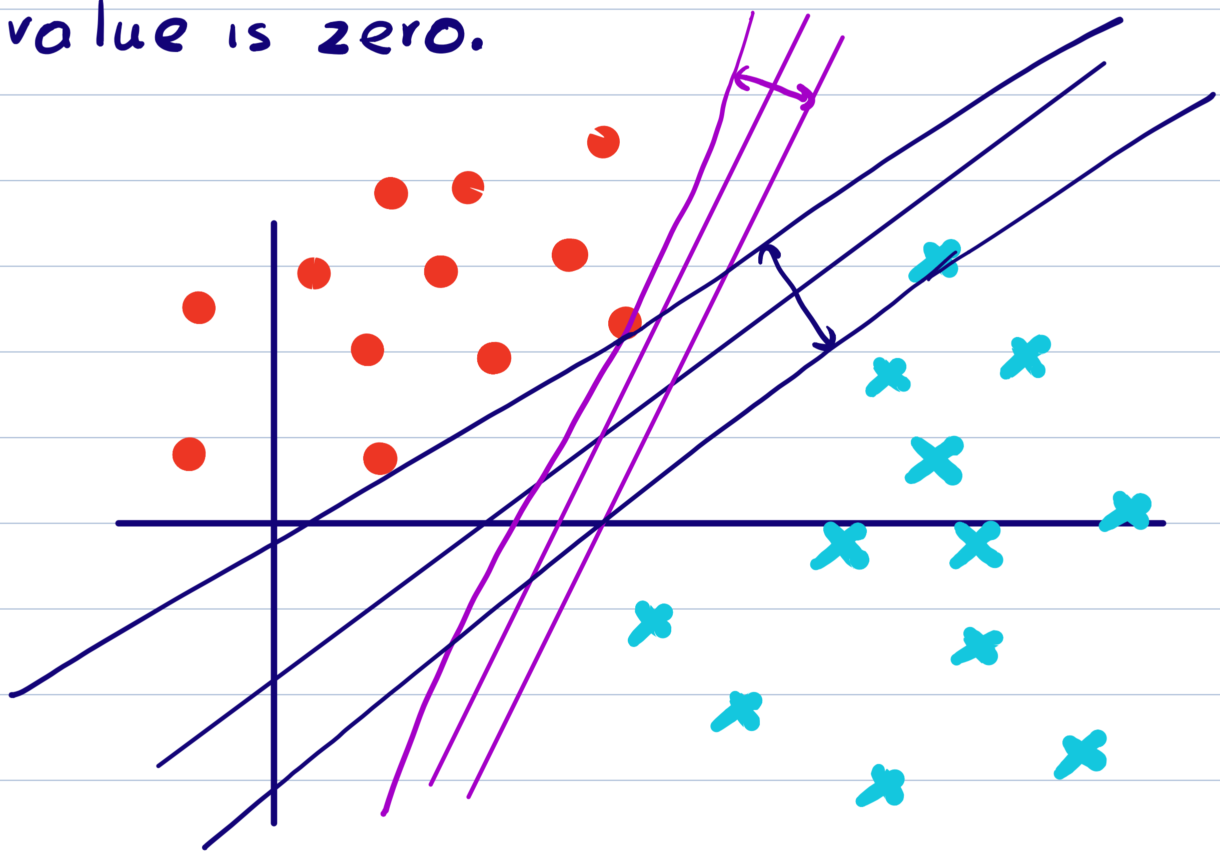
can be expressed as: $\quad \min_{z,\, a,\, b} \sum_{i=1}^{n} z_i$

$\quad$ st $\quad 0 \leq z_i$

$\quad\quad 1 - s_i(a^T x_i + b) \leq z_i$

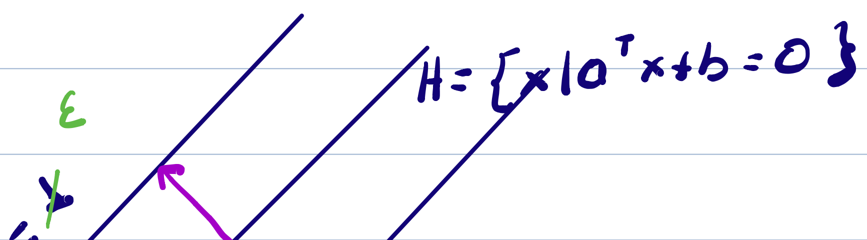
constants:
x_i, s_i

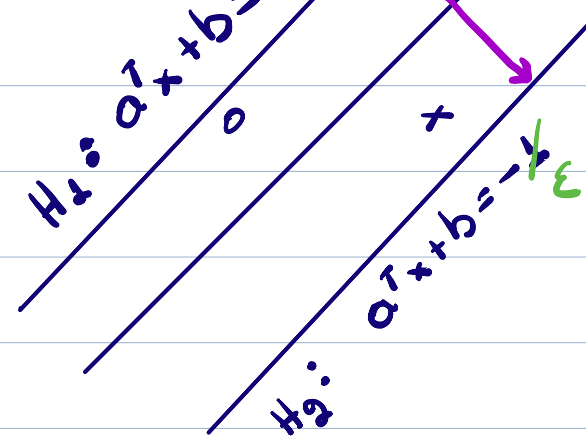
If the training data are linearly separable (there exists a hyperplane that perfectly separates them) the objective value is zero.



SVM (support vector machine):

maximize the "margin" around the separating hyperplanes





distance between H_1 & H_2 equals

$$\frac{2\epsilon}{\|a\|_2}$$

to max the margin we need to minimize $\|a\|$

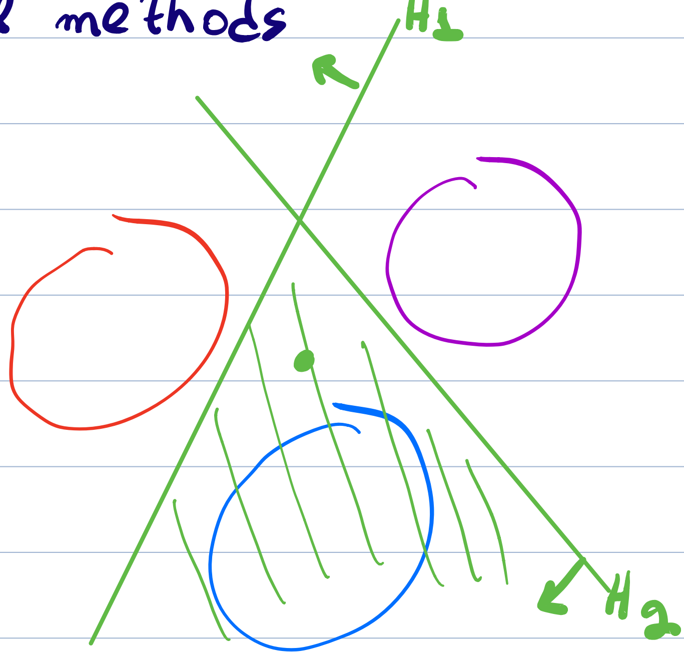
$$\min_{a,b} \left\{ \left(\sum \max\{0, 1 - s_i(a^T x_i + b)\} \right) + \lambda \|a\|_2^2 \right\}$$

constant weight

quadratic programming

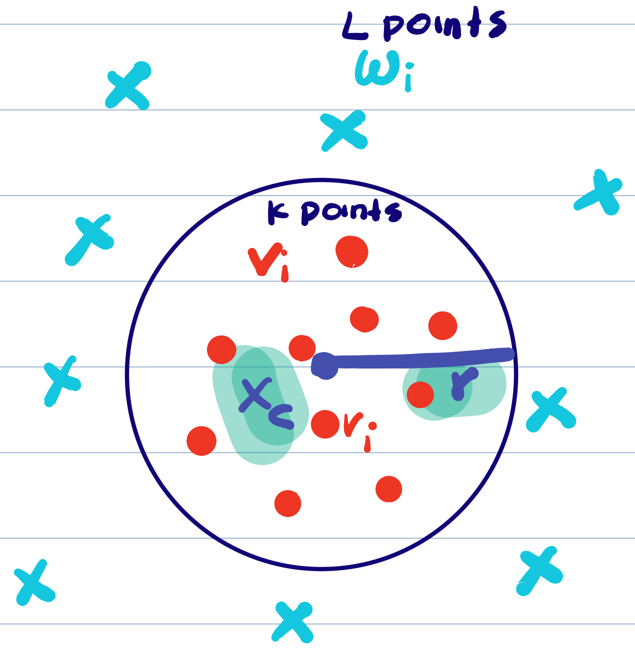
(Several other algorithms exist

- perceptron algorithm
- kernel methods)



Example 4

Find a sphere that separates two sets of points v_i and ω_i



Sphere: all points within distance r from a center x_c.

$$S = \{ x \mid \| x - x_c \|_2 \leq r \} =$$

$$= \{ x \mid (x - x_c)^T (x - x_c) \leq r^2 \}$$

Variables: r, x_c

Want for
$$\left.\begin{array}{ll} i = 1, \dots, K & (v_i - x_c)^T (v_i - x_c) \leq r^2 \\ i = 1, \dots, L & (\omega_i - x_c)^T (\omega_i - x_c) \geq r^2 \end{array}\right\} \Rightarrow$$

$$v_i^T v_i - 2 v_i^T x_c + x_c^T x_c \leq r^2$$

$$\omega_i^T \omega_i - 2 \omega_i^T x_c + x_c^T x_c \geq r^2$$

Change of variables: $x_c \to$ one variable

$\gamma \to r^2 - x_c^T x_c$

$\|v_i\|^2 - 2v_i^T x_c \leq \gamma \rightsquigarrow$ if not $\|v_i\|^2 - 2v_i^T x_c - \gamma > 0$

$\|w_i\|^2 - 2w_i^T x_c \geq \gamma \rightsquigarrow$ if not $\gamma - \|w_i\|^2 - 2w_i^T x_c$

$$\boxed{\min_{x_c, \gamma} \sum_{i=1}^{L} \max\{0, \gamma - \|w_i\|^2 - 2w_i^T x_c\} + \sum_{i=1}^{K} \max\{0, \|v_i\|^2 - 2v_i^T x_c - \gamma\}}$$

can be solved by solving an equivalent LP

Solving the LP

we retrieve x_c = center ✓

$\gamma = r^2 - \|x_c\|^2$ $\rightsquigarrow$ can we always solve this to find r?

How do we know that $\gamma + \|x_c\|^2$ is nonnegative?

→ There has to be at least one v_i inside sphere.

$$\gamma + \|x_c\|^2 \geq \|v_i\|^2 - 2v_i^T x_c + \|x_c\|^2 = \|v_i - x_c\|^2 \geq 0$$