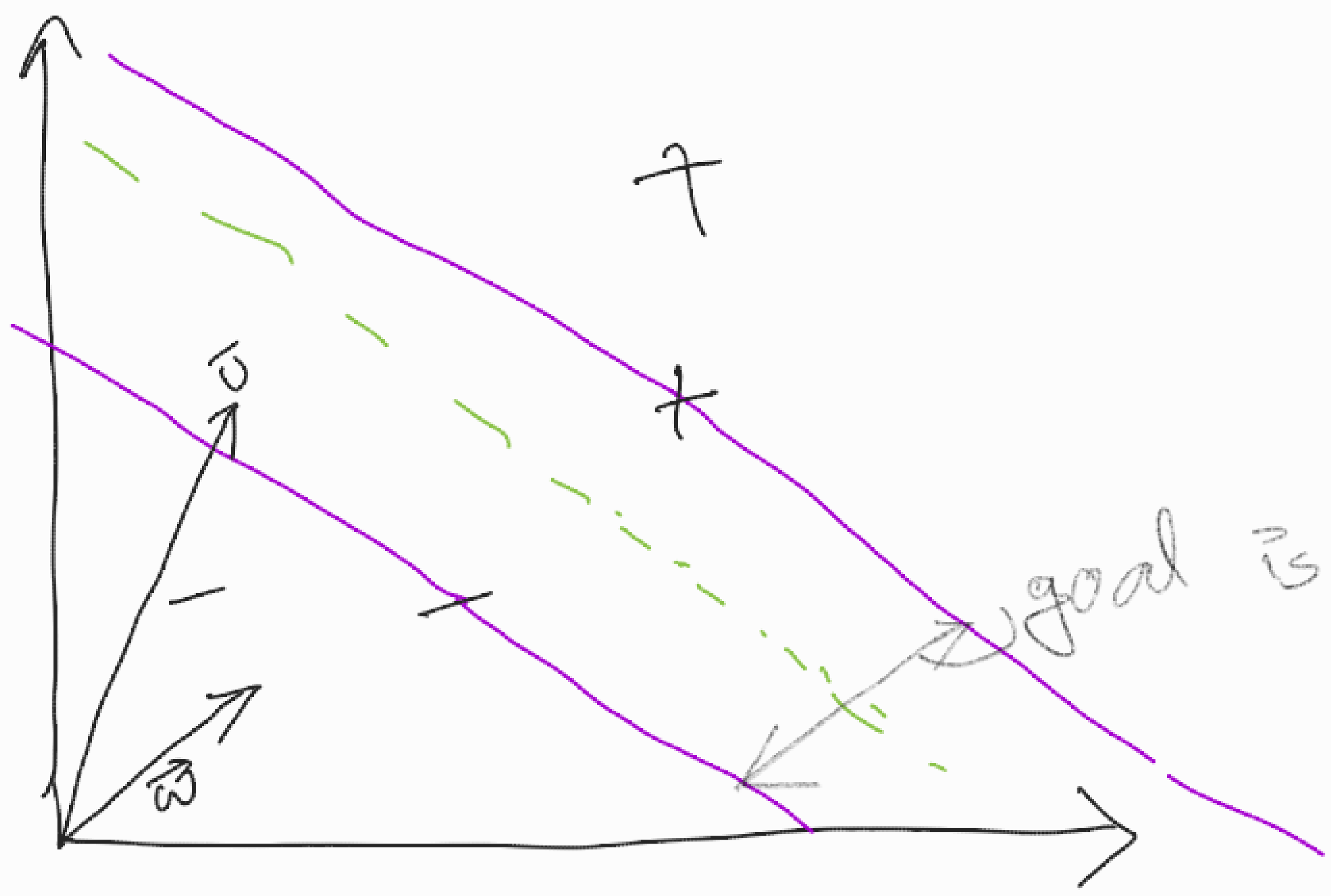
supervised learning problem that is used for classification purpose.



this algorithm tries to make the seperation street between positive & negetive example as wide as possible

$\vec{w}$ is a vector of any length & it is perpendicular to the median line of the street.

$\vec{u}$ ⇒ is the unkown vector that we want to classify

to classify $\vec{u}$:

$$\vec{w} \cdot \vec{u} \geq c \quad \text{—— (I)}$$

Here c is some constant. $\vec{w} \cdot \vec{u}$ gives the projection of $\vec{u}$ in the direction of $\vec{w}$. And c is the distance of the median line in the direction of $\vec{w}$

$$\boxed{\begin{array}{l} \vec{w} \cdot \vec{u} + b \geq 0 \quad \text{Then } + \\ \text{Decision rule} \end{array}}$$

Assumming $c = -b$

—— (II)

We don't know the value of b yet. we also don't know the length of w. There isn't enough constraint to calculate $\vec{w}$ & b. Now we will put some additional constraint so that we can calculate $\vec{w}$ & b.

$$\vec{w}\cdot\vec{x}_{+}+b \geq 1$$
$$\vec{w}\cdot\vec{x}_{-}+b \leq -1$$

this gives a seperation between two classification.
This implies all unit positive sample must be 1 unit distance from the decision boundary

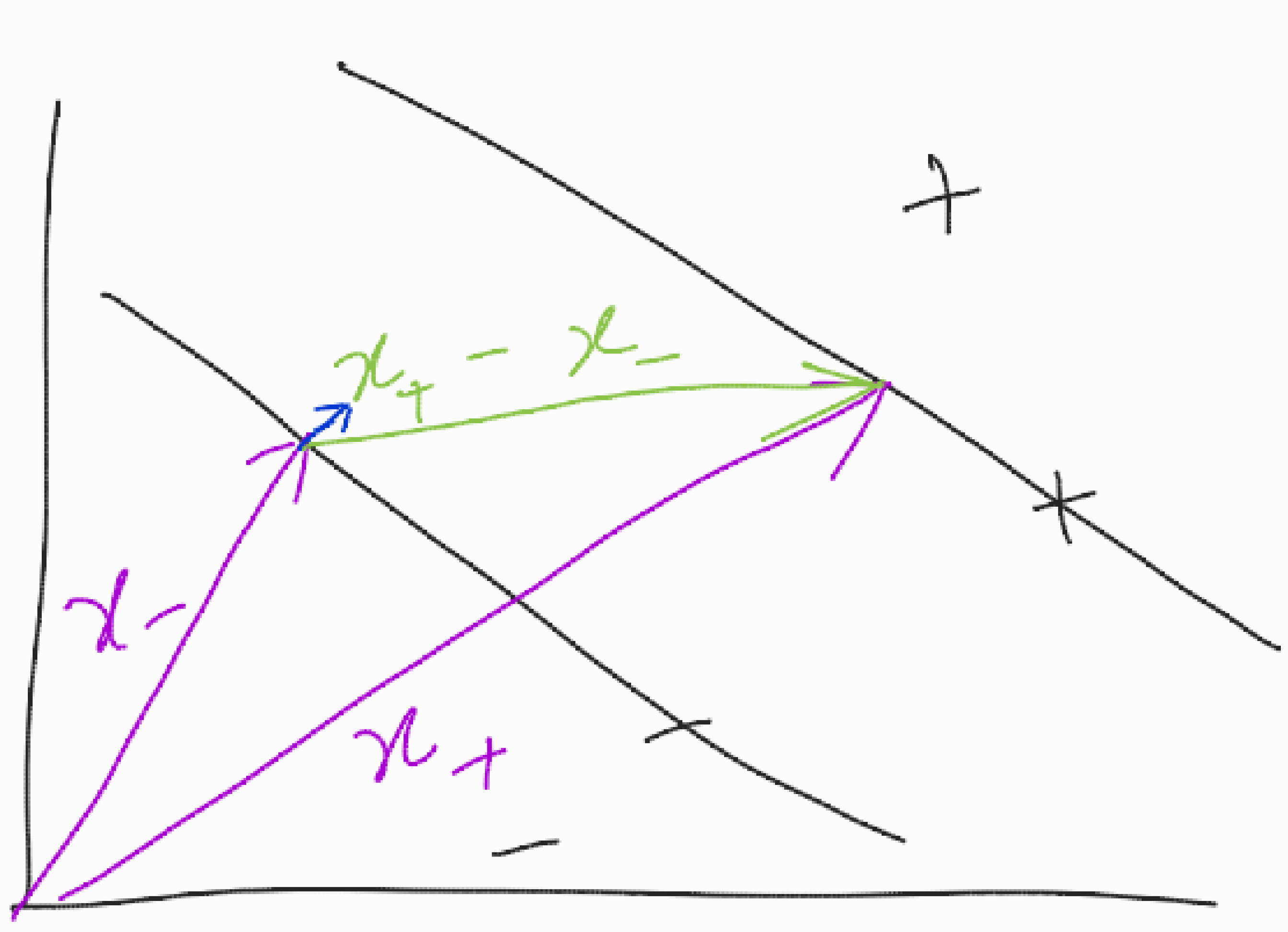
we will combine these two equation into one.

$$y^{(i)} = \begin{cases} +1 & \text{for positive sample.} \\ -1 & \text{for negetive sample.} \end{cases}$$

if we combine above two equation we get:-

$$y^{(i)}(\vec{w}\cdot\vec{x}^{(i)}+b) \geq 1$$
$$y^{(i)}(\vec{w}\vec{x}^{(i)}+b) - 1 \geq 0.$$

$$\boxed{y^{(i)}(\vec{w}\vec{x}^{(i)}+b) - 1 = 0} \quad \text{for } x^{(i)} \text{ in the gutter.} \rightarrow (3)$$



$x_+ - x_-$ gives width of the street. But its direction is not correct. The direction should be perpendicular of the median line. We already know $\vec{w}$ is a vector that is perpendicular to the median line. So if we multiply unit vector of $\vec{w}$ with $x_+ - x_-$ then we will get the width of the street.

$$\text{width} = (\vec{x}_+ - \vec{x}_-)\frac{\vec{w}}{\|\vec{w}\|}$$

$$= (\vec{x}_+ \vec{w} - \vec{x}_- . \vec{w})\frac{1}{\|\vec{w}\|}$$

$$= (1 - b + 1 + b)\frac{1}{\|\vec{w}\|}$$

$$= \frac{2}{\|\vec{w}\|} \quad \text{— (4)}$$

From (2).
for +:-
$wx + b - 1 = 0.$
$wx = 1 - b.$
for -:
$-wx - b - 1 = 0.$
$wx = -1 - b.$

So we want to maximize the width of the street

$$\text{Max} \frac{2}{\|\vec{w}\|}$$

$$\Rightarrow \text{Max} \frac{1}{\|\vec{w}\|}$$

$$\Rightarrow \text{Min} \|\vec{w}\| \quad \leftarrow \text{non convex problem}$$

$$\Rightarrow \text{Min} \|\vec{w}\|^2. \quad \leftarrow \text{convex problem}$$

$$\boxed{\text{Min } \frac{1}{2}\|\vec{w}\|^2} \quad (5)$$

$\frac{1}{2}$ is for mathematical convenience.

(5) is a function that we want to minimize while we need to maintain the constraint (3).

We are going to use Lagrange for this purpose.

$$\mathcal{L} = \frac{1}{2}\|\vec{w}\|^2 - \sum_i \alpha^{(i)}\left[y^{(i)}(\vec{w}\vec{x}^{(i)} + b) - 1\right] \quad (6)$$

According to Lagrange

$$\nabla \mathcal{L} = 0 \quad \Rightarrow \quad \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \vec{w}} \\ \frac{\partial \mathcal{L}}{\partial b} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\frac{\partial \mathcal{L}}{\partial \vec{w}} = \vec{w} - \sum \alpha^{(i)} y^{(i)} \vec{x}^{(i)} = 0$$

$$\Rightarrow \boxed{\vec{w} = \sum_{i=1}^{m} \alpha^{(i)} y^{(i)} \vec{x}^{(i)}} \quad (7)$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_i \alpha^{(i)} y^{(i)} = 0$$

$$\Rightarrow \boxed{\sum_i \alpha^{(i)} y^{(i)} = 0} \quad (8)$$

putting these value in lagrang equation.

$$\mathcal{L} = \frac{1}{2}\left(\sum_i \alpha^{(i)} y^{(i)} \vec{x}^{(i)}\right) \cdot \left(\sum_j \alpha^{(j)} y^{(j)} \vec{x}^{(j)}\right) - \left(\sum \alpha^{(i)} y^{(i)} \vec{x}^{(i)}\right) \cdot \left(\sum \alpha^{(j)} y^{(j)} \vec{x}^{(j)}\right)$$
$$- b\sum \alpha^{(i)} y^{(i)} + \sum \alpha^{(i)}$$

$$= \sum_i \alpha^{(i)} - \frac{1}{2} \sum_i \sum_j \alpha^{(i)} \alpha^{(j)} y^{(i)} y^{(j)} \boxed{\vec{x}^{(i)} \vec{x}^{(j)}}$$

⑨

maximization only depends on these dot product.

From equation ① the decision rule -

$$\underbrace{\sum \alpha^{(i)} y^{(i)} \vec{x}^{(i)}}_{} \cdot \vec{u} + b > 0 \quad \text{THEN } +$$

$$\Rightarrow \vec{w} \cdot \vec{u} + b > 0 \quad \text{THEN } +.$$

Coding :-

In training steps we need to find the weights for each of the input feature.

From ⑦ we know, $\vec{w} = \sum_{i=1}^{m} \alpha^{(i)} x^{(i)} y^{(i)}$

from the training data we already know $x^{(i)}, y^{(i)}$ so we need to find $\alpha^{(i)}$. For that we have lagrange maximization function ⑨ (find $\alpha^{(i)}$ that will maximize L)

$$L = \sum_{i=1}^{m} \alpha^{(i)} - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha^{(i)} \alpha^{(j)} y^{(i)} y^{(j)} \vec{x}^{(i)} \cdot \vec{x}^{(j)}.$$

we also have to maintain the constraint

$$\alpha^{(i)} \geq 0$$

And from ⑧ $\sum^{m} \alpha^{(i)} y^{(i)} = 0$

We will use python package cvxopt to solve this quadratic optimization problem. This package only takes vector. so we need to convert our file according to the format of this package.

$$L = \sum_{i}^{n} \alpha_i - \frac{1}{2}\left(\sum_{i}^{n} \alpha_i y_i x_i\right) \cdot \left(\sum_{i}^{n} \alpha_j y_j x_j\right)$$

$$L = 1^T \alpha - \frac{1}{2}(\alpha^T X) \cdot (\alpha^T X) \qquad \text{since } X = [x_1 y_1 ...]$$

$$L = 1^T \alpha - \frac{1}{2}(\alpha^T X)^T (\alpha^T X)$$

$$L = 1^T \alpha - \frac{1}{2}\alpha X^T \alpha^T X$$

$$L = 1^T \alpha - \frac{1}{2} X^T \alpha^T X \alpha \longleftarrow \text{we are maximizing this.}$$

Instead of maximizing $\mathcal{L}$ we can minimize $-\mathcal{L}$

$$\min_{\alpha} \mathcal{L} = -\frac{1}{2} X^T \alpha^T X \alpha - 1^T \alpha.$$

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array} \longleftarrow$$

SVM Problem

$$max_{\boldsymbol{\alpha}}\ \mathbf{1}^T\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{\alpha}$$

Subject to:

$$\boldsymbol{\alpha} \geqslant \mathbf{0}$$
$$\boldsymbol{y}^T \boldsymbol{\alpha} = 0$$

$$X = \alpha$$

$$P = X^T X$$

$$q = -1_N$$

$$G = -1_{NxN}$$

$$h = 0_N$$

$$A = y^T$$

$$v = 0$$