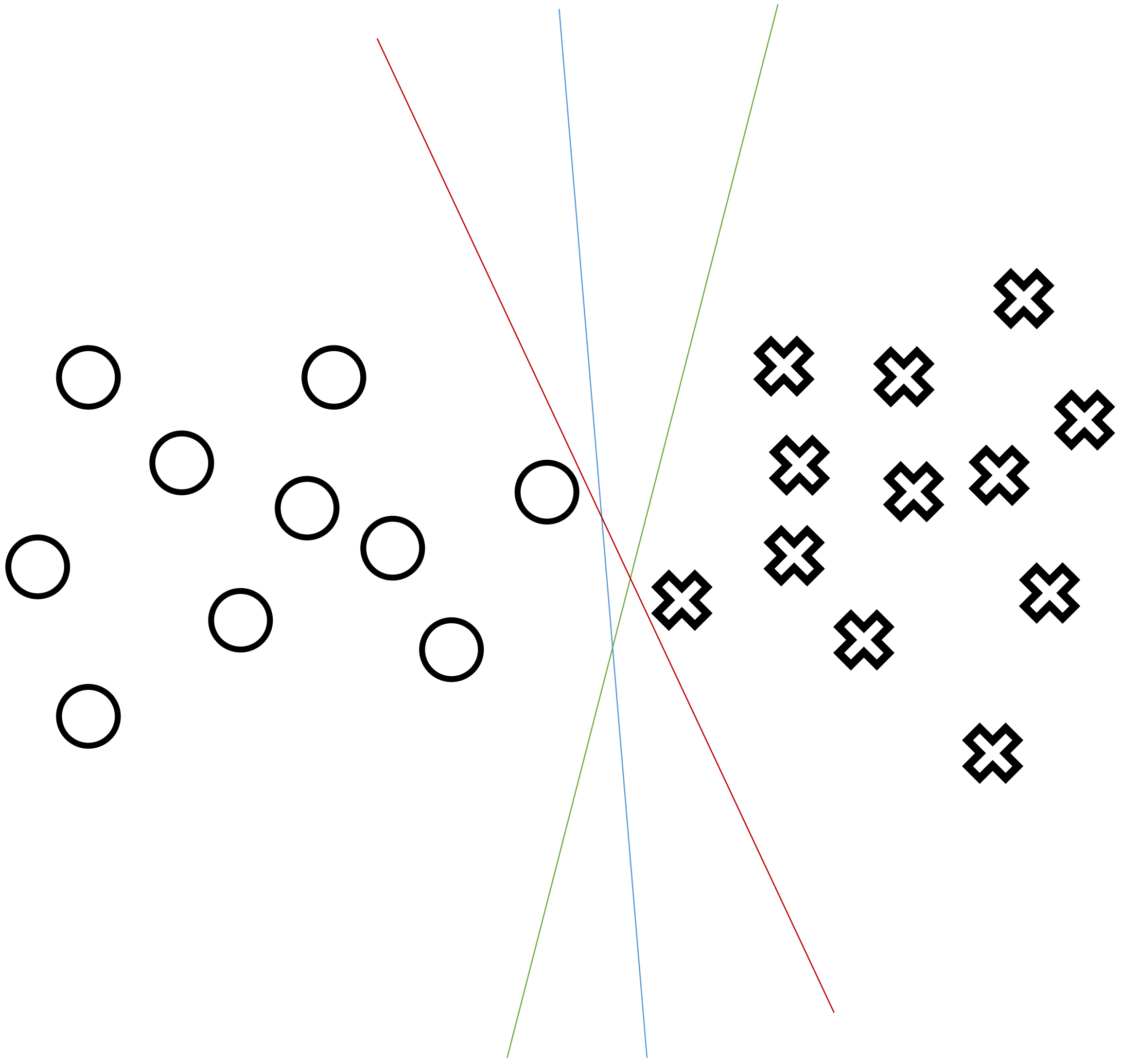
Support Vector Machine SVM

JrPhy

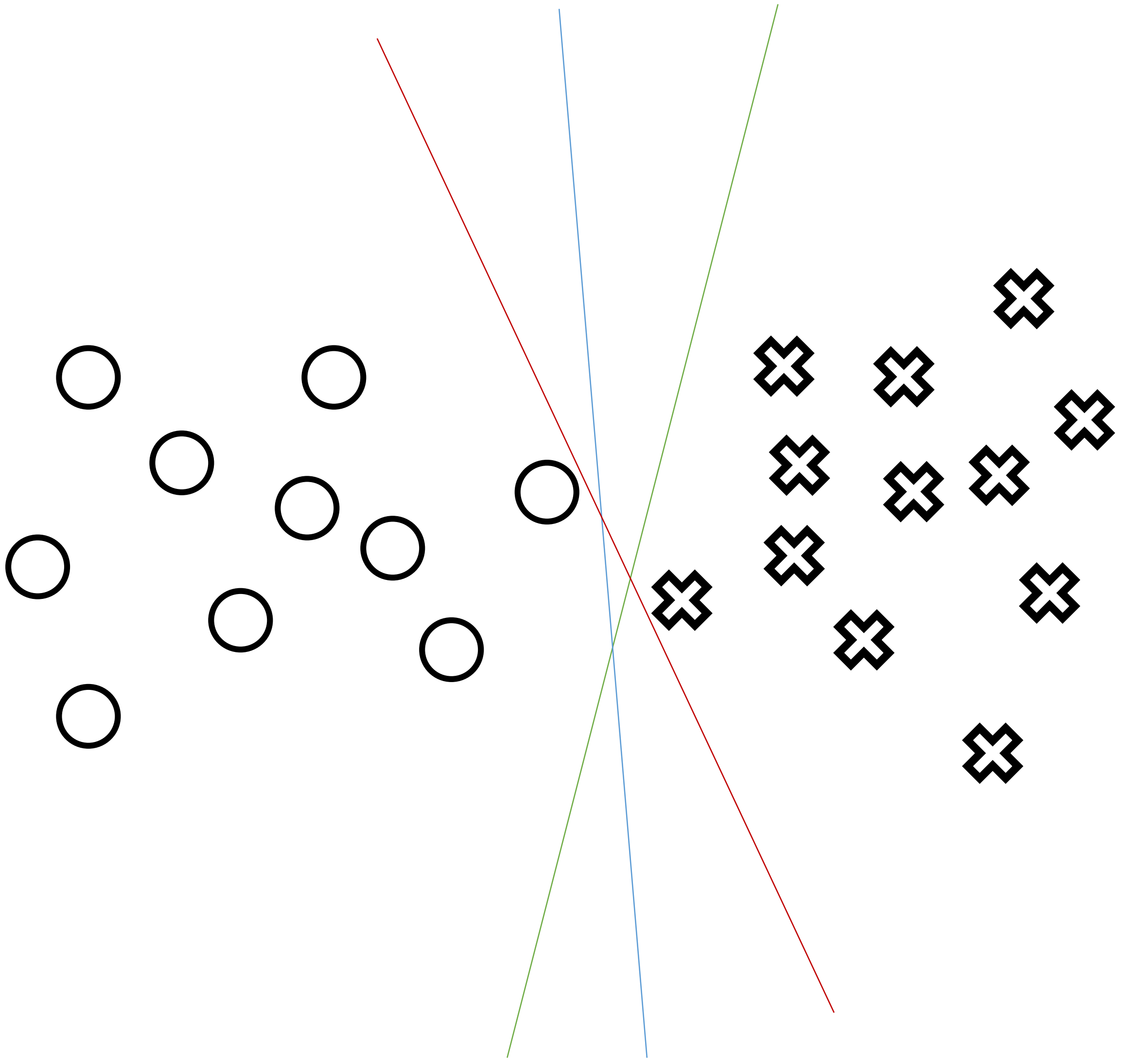
Introduction

- In machine learning, the basic problem is classification, there are many methods to classify the binary problem. PLA is one of the solution, but its solution is not unique, so we add some constrains to get a unique line or hyperplane.



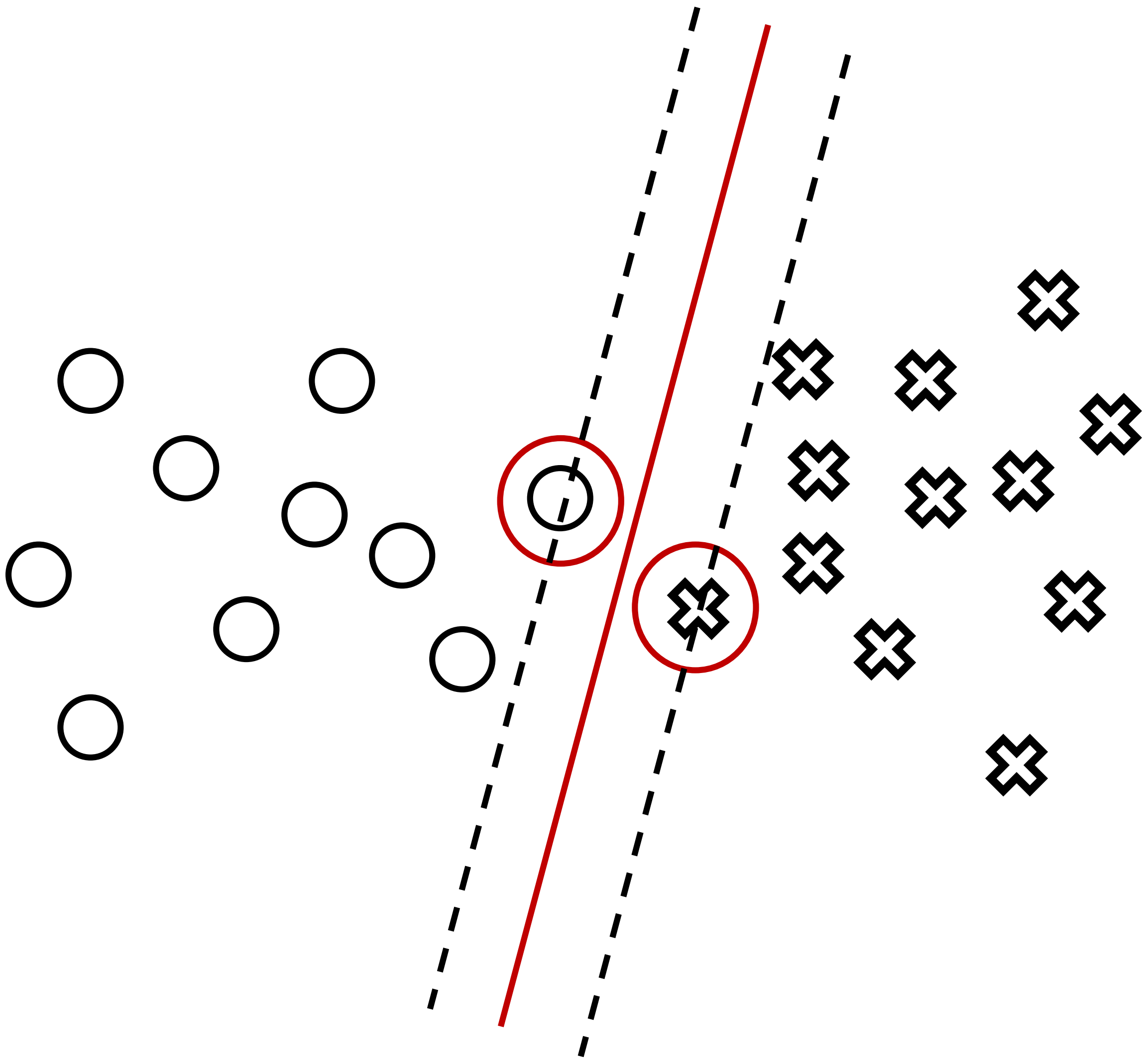
Constrain

- The constrain is the distant of the closest data is the largest
- There are 3 lines, all are separate “O” and “X”. the red line is closest to the data, and the blue is in the middle, the green line is the farthest to the data. So the green line is our candidate in this example.



Constrain

- In this dataset, we just care about 2 points, and the red line is farthest from the closest point, so we want this line.
- Then like PLA, the line eq. is
- $<\boldsymbol{w}, \boldsymbol{x}^{\boldsymbol{T}}> = \begin{bmatrix} w_2 & w_1 & w_0 \end{bmatrix} \begin{bmatrix} x_2 \\ x_1 \\ x_0 \end{bmatrix}$
- $= w_2 x_2 + w_1 x_1 + w_0 x_0 = 0$
- $w_0 x_0$ is the constant term

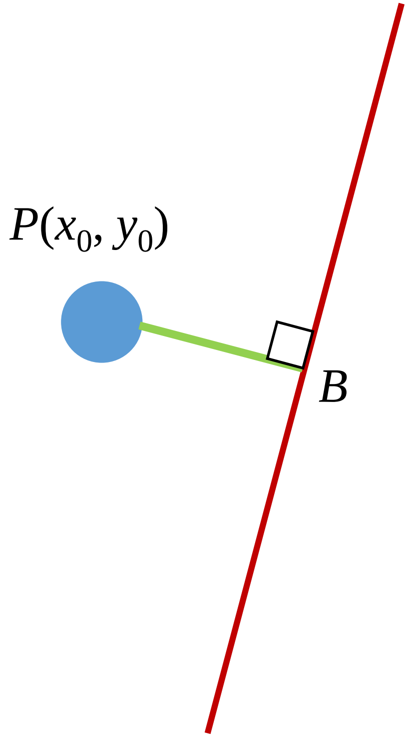


Derivation

- The definition of distance from a point to a line means the perpendicular distance in Euclidian space, it's also the "shortest distance". So there exists only one distance.
- Point $P(x_0, y_0)$ to a line L: $ax+by+c = 0$ is $d(p, L)$

$$d(P,L) = \frac{\left|ax_0 + by_0 + c\right|}{\sqrt{a^2 + b^2}}$$

- Suppose L' passes through P and perpendicular to L, then
- L' : $bx - ay = (bx_0 - ay_0)$, L: $ax+by+c = 0$
- Solve x, y, we get $$B(x,y) = \left(\frac{b^2x_0 - aby_0 - ac}{a^2 + b^2}, \frac{a^2y_0 - abx_0 - bc}{a^2 + b^2}\right)$$



Distance

- By the definition of two point, then we get $d(P, L) = \overline{PB}$

$$d(P,L)=\sqrt{\left(x_0-\frac{b^2x_0-aby_0-ac}{a^2+b^2}\right)^2+\left(y_0-\frac{a^2y_0-abx_0-bc}{a^2+b^2}\right)^2}$$

$$=\frac{1}{a^2+b^2}\sqrt{\left(a^2x_0+aby_0+ac\right)^2+\left(b^2y_0+abx_0+bc\right)^2}$$

$$=\frac{1}{a^2+b^2}\sqrt{(a^2+b^2)c^2+a^2x_0{}^2(a^2+b^2)+b^2y_0{}^2(a^2+b^2)}$$

$$\boxed{=\frac{\left|ax_0+by_0+c\right|}{\sqrt{a^2+b^2}}}$$

Distance

- Let's observe, $<\boldsymbol{w}, \boldsymbol{x}^T> = w_2x_2+w_1x_1+w_0x_0 = ax+by+c$
- So that $\boldsymbol{w} = (a, b)$ is a vector, $\boldsymbol{x} = (x_0, y_0)$ is a vector, $\| \boldsymbol{w} \| = \sqrt{a^2+b^2}$, $w_0x_0 = c = b$,

$$d(P,L) = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2+b^2}} = \frac{|w^T x + b|}{\|w\|}$$

- $<\boldsymbol{w}, \boldsymbol{x}^T>$ I write as $\boldsymbol{w}^T\boldsymbol{x}$, means the inner product.
- The others are like PLA, y is +1 or -1
- $y(\boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{b}) > 0 \rightarrow$ correct, $y(\boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{b}) < 0 \rightarrow$ incorrect
- Then want to optimize the distance.

Optimize

- Combine the label → $y(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + \boldsymbol{b}) = 1$. But we don't need so strong condition, just need
  - $y(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + \boldsymbol{b}) \geq 1$
- After this condition is determined, then want to find the maxima distance, so

$$\max d(P, L) = \max \frac{\left| w^T x + b \right|}{\left\| w \right\|}$$

- By previous condition, max $d(P, L)$ is equivalent to

$$\max \frac{1}{\left\| w \right\|}$$

Optimize

$$\max_{b,w} \frac{1}{\|w\|} = \min_{b,w} \|w\| = \min_{b,w} \sqrt{w^2} \sim \min_{b,w} \frac{1}{2}\|w\|^2$$

- Here we want to minimize <$\boldsymbol{w^T}$, $\boldsymbol{w}$> with constrain $y(\boldsymbol{w^T x} + \boldsymbol{b}) \geq 1$, we can use "Lagrange undetermined multiplier" to solve it.

$$L(\alpha, w, b) = \frac{1}{2}\|w\|^2 - \sum_i \alpha_i [y_i (w^T x + b) - 1]$$

- α is the undetermined multiplier, the 2nd term is bigger, then the 1st term smaller, so we want to solve

$$\boxed{\min_{b,w} \max_{\forall \alpha_i \geq 0} L(\alpha, w, b) = \max_{\forall \alpha_i \geq 0} \min_{b,w} L(\alpha, w, b)}$$

- We can use KKT condition to change the order.

Optimize

$$\frac{\partial}{\partial w}L(\alpha,w,b)=\|w\|-\sum_i \alpha_i y_i x=0 \qquad \frac{\partial}{\partial b}L(\alpha,w,b)=\sum_i \alpha_i y_i =0$$

- Substitute to the $L(\alpha, w, b)$
- $$L(\alpha,w,b)=\frac{1}{2}\|w\|^2-\sum_i \alpha_i[y_i(w^T\vec{x}+b)-1]$$

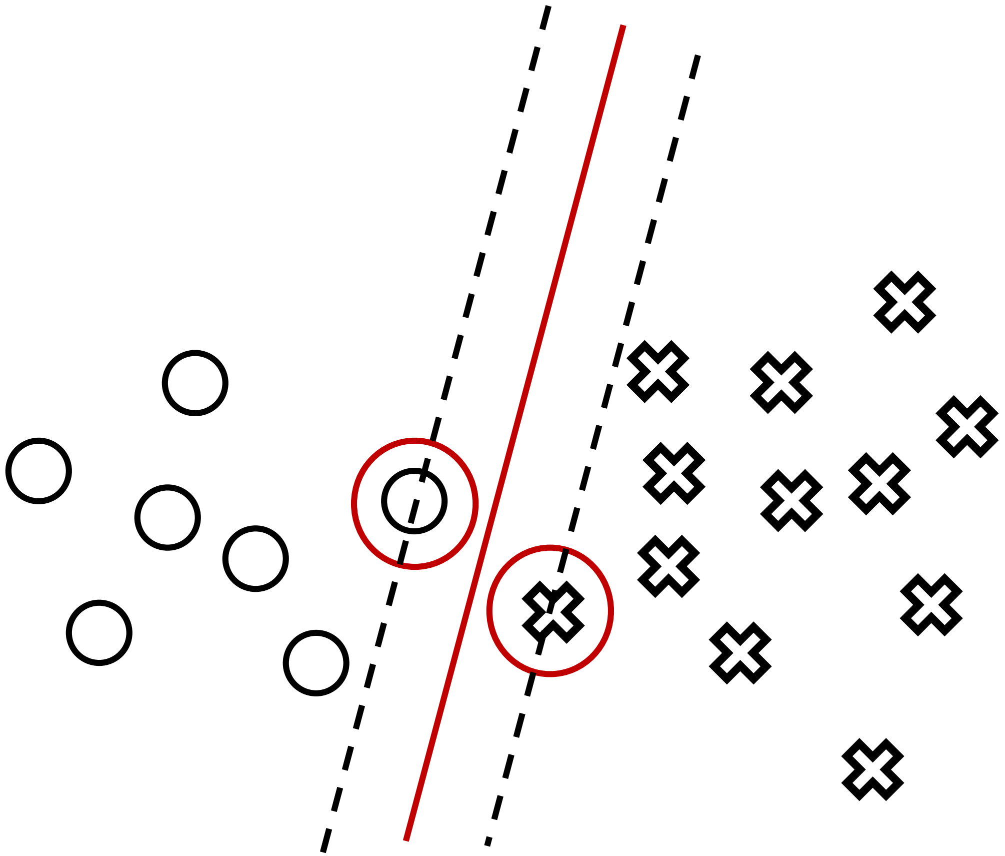
$\boldsymbol{ww^T}$ $\qquad$ 0

$$=\frac{1}{2}\left\|\sum_i \alpha_i y_i \vec{x}\right\|^2-\sum_i(\alpha_i y_i w^T\vec{x}+\alpha_i y_i b-\alpha_i)$$

$$=-\frac{1}{2}\left\|\sum_j\sum_i \alpha_i\alpha_j y_i y_j \vec{x}^T\vec{x}'\right\|^2+\sum_i \alpha_i$$

Optimize

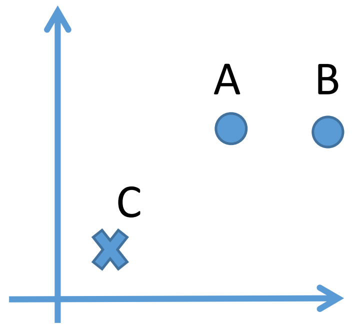
- Here we have determined the optimal *w*, next we want to solve optimal *b*. In the 2nd term, if $\alpha = 0$, *b* can be any number, so here we want to solve $\alpha > 0$.
- $\alpha[y(\mathbf{w}^T\mathbf{x} + \mathbf{b}) - 1] = 0 \rightarrow y(\mathbf{w}^T\mathbf{x} + \mathbf{b}) = 1$
- This means we get $\alpha > 0$, these points are on the boundary, we call it **support vector**.



$$\frac{\partial}{\partial b} L(\alpha, w, b) = \boxed{\sum_i \alpha_i y_i = 0}$$

$$\max_{\alpha \geq 0} L(\alpha, w, b) = \max_{\alpha \geq 0} \left(-\frac{1}{2} \left\| \sum_j \sum_i \alpha_i \alpha_j y_i y_j \vec{x}^T \vec{x}' \right\|^2 + \sum_i \alpha_i \right) = \boxed{\min_{\alpha \geq 0} \left(\frac{1}{2} \left\| \sum_j \sum_i \alpha_i \alpha_j y_i y_j \vec{x}^T \vec{x}' \right\|^2 - \sum_i \alpha_i \right)}$$

Example



- A(3, 3, 1), B(4, 3, 1), C(1, 1, -1)
- The last is label.

$$\sum_i \alpha_i y_i = 0 \rightarrow \alpha_1 + \alpha_2 = \alpha_3$$

$$\min_{\forall \alpha_i \geq 0} \left(\frac{1}{2} \left\| \sum_j \sum_i \alpha_i \alpha_j y_i y_j x^T x' \right\|^2 - \sum_i \alpha_i \right) = \min_{\forall \alpha_i \geq 0} \left[\frac{1}{2} \left(8\alpha_1^2 + 20\alpha_1\alpha_2 + 13\alpha_2^2 \right) - 2\alpha_1 - 2\alpha_2 \right]$$

$$\frac{\partial}{\partial \alpha_1} \left[\frac{1}{2} \left(8\alpha_1^2 + 20\alpha_1\alpha_2 + 13\alpha_2^2 \right) - 2\alpha_1 - 2\alpha_2 \right] = 8\alpha_1 + 10\alpha_2 - 2 = 0$$

$$\frac{\partial}{\partial \alpha_3} \left[\frac{1}{2} \left(8\alpha_1^2 + 20\alpha_1\alpha_2 + 13\alpha_2^2 \right) - 2\alpha_1 - 2\alpha_2 \right] = 10\alpha_1 + 13\alpha_2 - 2 = 0$$

$$\alpha_1 = 1.5$$

$$\alpha_2 = -1$$

Example

- The solution does not satisfy the constrain, so the minimum is on the boundary.

$$\alpha_1 = 0 \rightarrow \frac{13}{2}\alpha_3^{\ 2} - 2\alpha_3 = \frac{13}{2}\left(\alpha_3 - \frac{2}{13}\right)^2 - \frac{2}{13} \rightarrow \alpha_3 = \alpha_2 = \frac{2}{13}, \min = -\frac{2}{13}$$

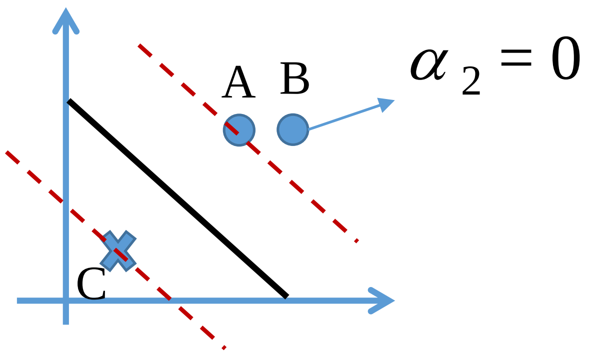
$$\alpha_2 = 0 \rightarrow 4\alpha_1^{\ 2} - 2\alpha_1 = 4\left(\alpha_1 - \frac{1}{4}\right)^2 - \frac{1}{4} \rightarrow \alpha_1 = \alpha_3 = \frac{1}{4}, \min = -\frac{1}{4}$$

The minimum

- The eq. of plane is $0.5x_1+0.5x_2 - 2 = 0$

$$w = \sum_i \alpha_i y_i \vec{x}_i = \frac{1}{4}\times(-1)\times(3,3) + \frac{1}{4}\times(-1)\times(4,3) = (\frac{1}{2}, \frac{1}{2})$$

$$b = y - w^T x = 1 - \frac{1}{4}\times(-1)\times 18 + \frac{1}{4}\times(1)\times 20 = 2$$



Summary

- In the example we can see that if $\alpha_i = 0$, then it's not on the boundary, and don't need to calculate, either.
- What we care about the data is on the boundary, that is, if there are more data outside the area, the solution is the same, too.