法律声明

关注 小象学院

Machine Learning

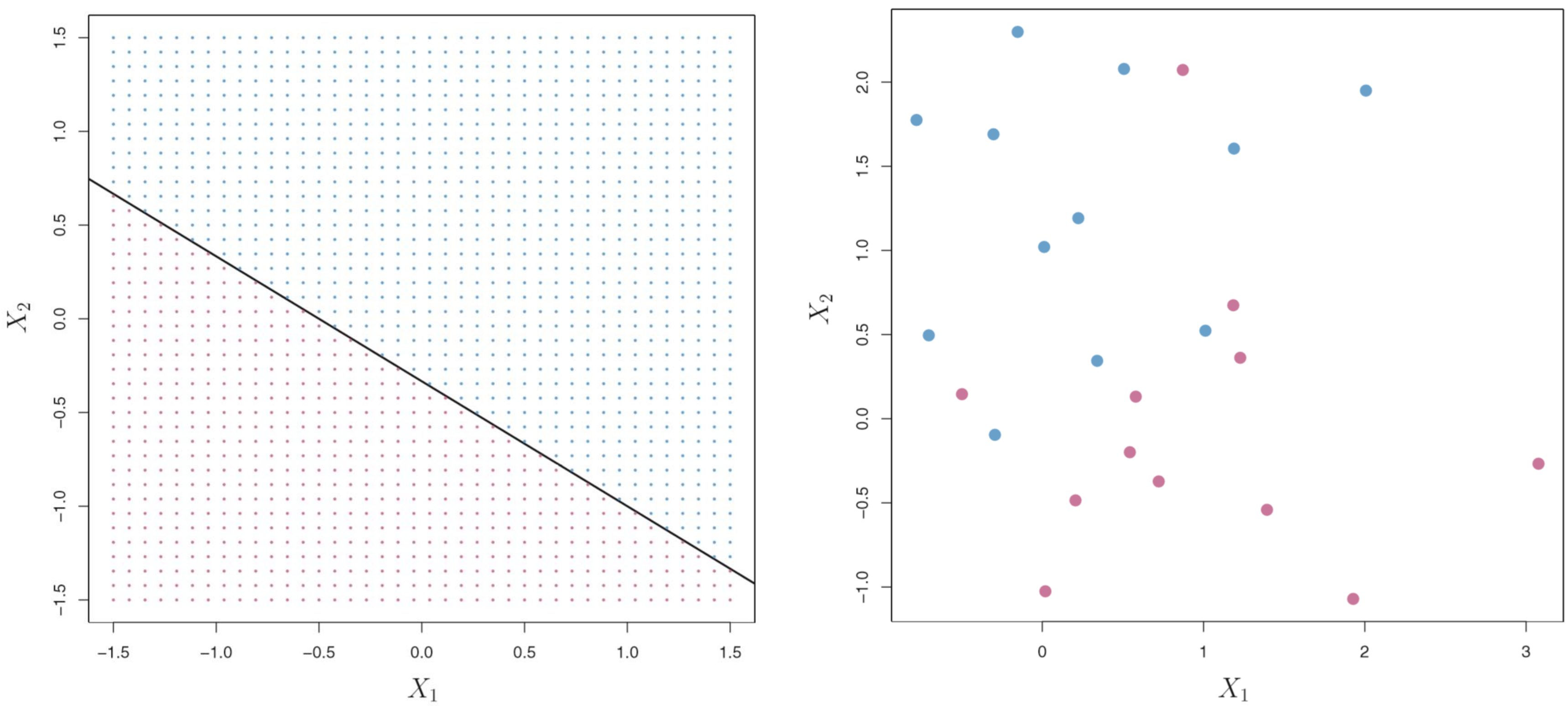
Part 4: Max Margin

Zengchang Qin (Ph.D.)

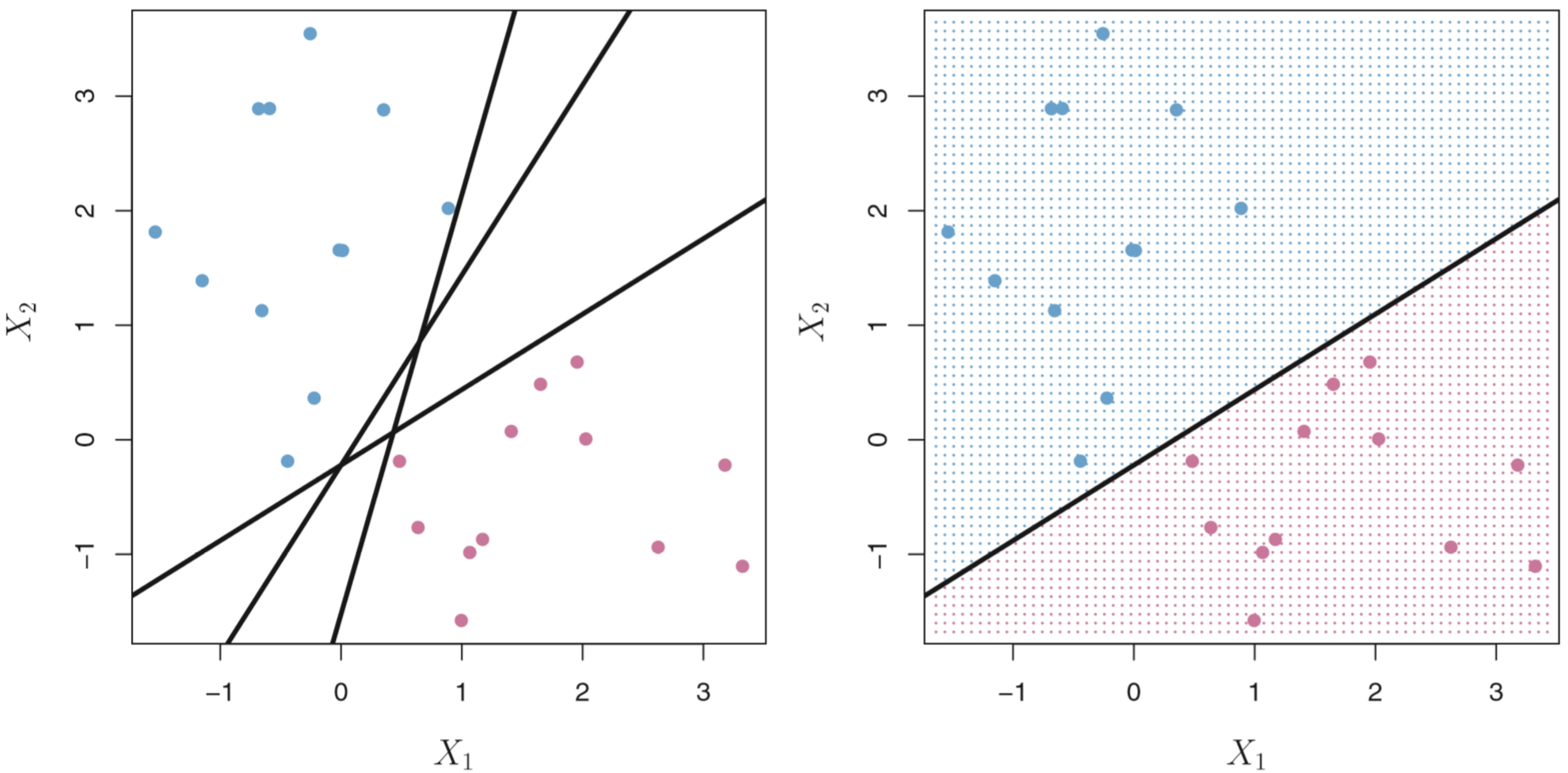
Classification with Max Margin

Hyperplane

The hyperplane 1+2X1+3X2 =0 is shown. The blue region is the set of points for which 1+2X1 +3X2 > 0, and the purple region is the set of points for which 1+2X1 +3X2 < 0.



Which One is Better?



The Problem

Given a data set,

$$D = \{ (\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \dots, (\boldsymbol{x_m}, y_m)\}, y_i \in \{-1,1\},$$

How can we find a hyperplane to classify them?
Is it the best one?

A hyperplane can be described as the following function

$$\boldsymbol{\omega}^{\boldsymbol{T}} \boldsymbol{x} + b = 0,$$

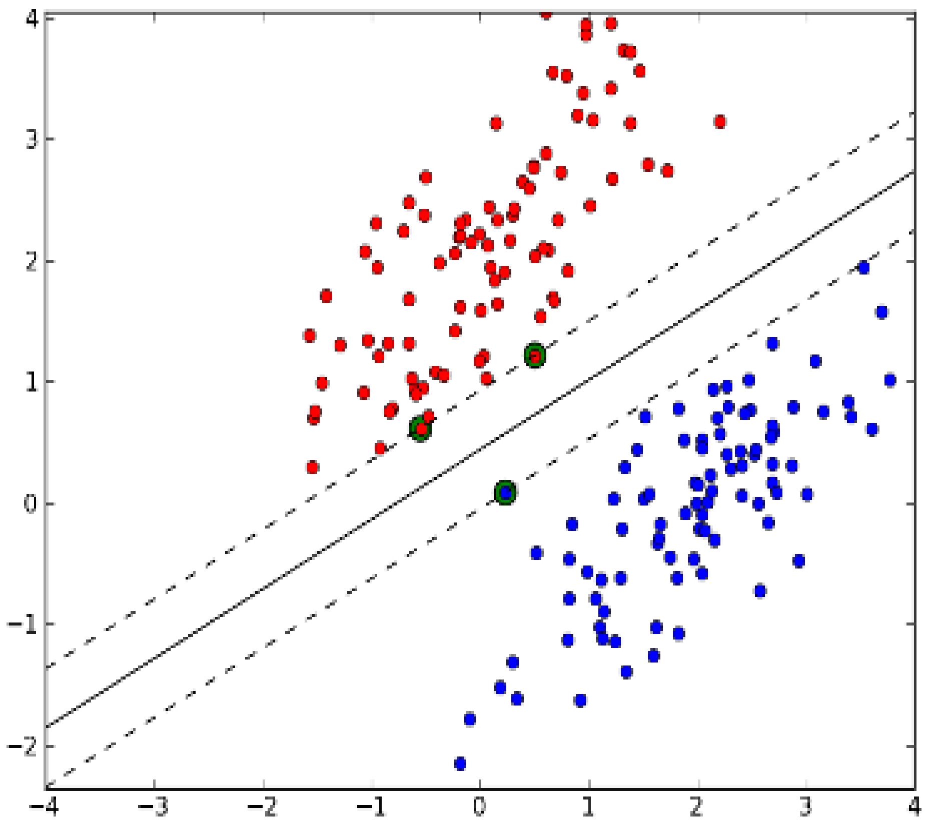
where $\boldsymbol{\omega} = \{\omega_{\boldsymbol{1}}; \omega_{\boldsymbol{2}}; \dots; \omega_{\boldsymbol{d}}\}$ is the normal vector of the hyperplane

$$\boldsymbol{\omega}^{\boldsymbol{T}} \boldsymbol{x_i} + b \geq +\sigma,$$

$$\boldsymbol{\omega}^{\boldsymbol{T}} \boldsymbol{x_i} + b \leq -\sigma,$$

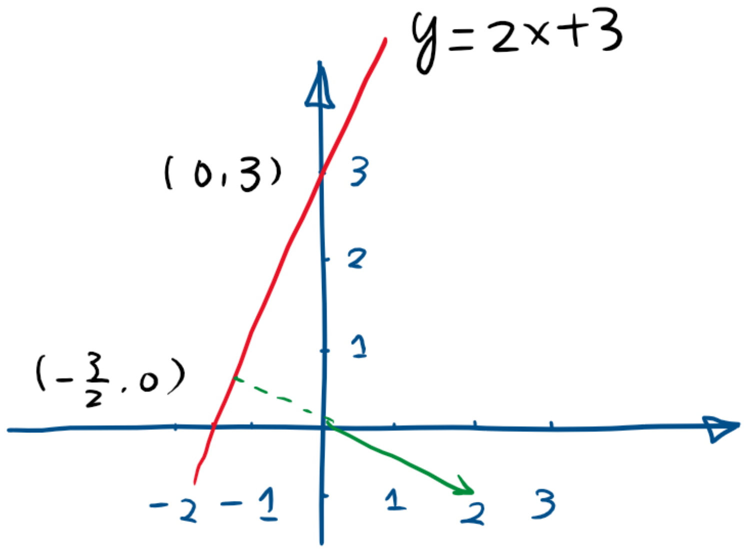
Normalize to $y_{\boldsymbol{i}}(\boldsymbol{\omega}^{\boldsymbol{T}} \boldsymbol{x_i} + b) >= 1, \quad i = 1,2, \dots, m.$

Support Vectors



Linear Function

1) Linear Function



Given a linear function $y = 2x+3$, we have $(0,3)$ and $(-\frac{3}{2}, 0)$ on the hyperplane.

$$y = 2x+3 \Rightarrow$$

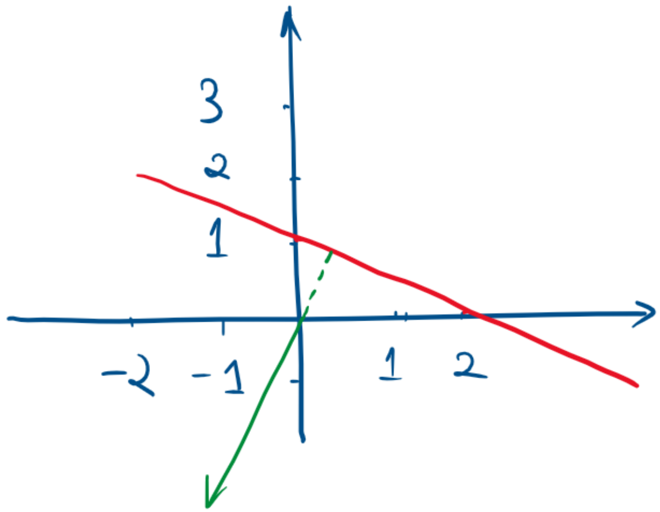
$$2x - y + 3 = 0$$

$$(2,-1)\begin{pmatrix} x \\ y \end{pmatrix} + 3 = 0$$

Probabilistic Interpretation

For a linear function $w^T x + b = 0$, the direction of w^T is perpendicular to the original linear function.

E.g.



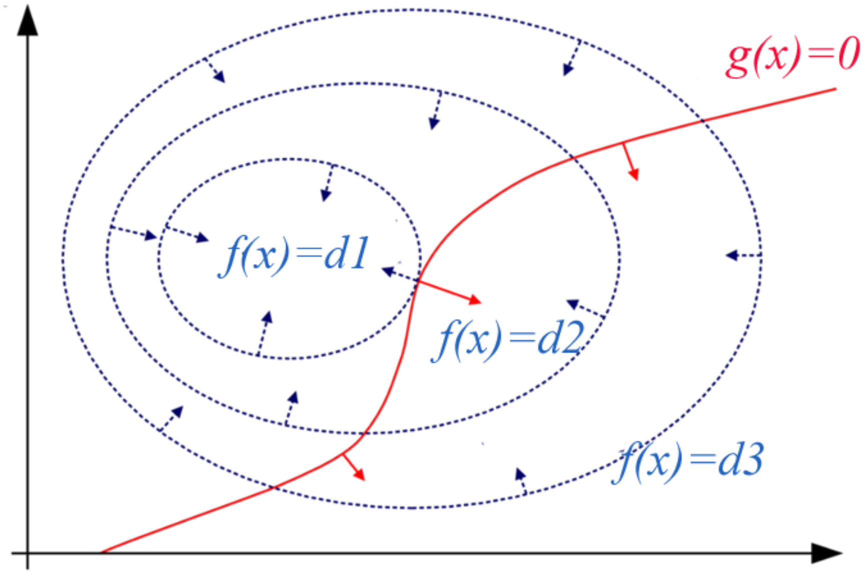
$y = -\frac{1}{2}x + 1$

$$\left(-\frac{1}{2}, -1\right)\begin{pmatrix} x \\ y \end{pmatrix} + 1 = 0$$

$$w^T = \left(-\frac{1}{2}, -1\right)$$

Lagrange Multiplier

Lagrange multipliers:



Case1: equality constraint

$$\min f(\boldsymbol{x})$$

$$s.t.: \quad g(\boldsymbol{x}) = 0$$

$$\nabla f(\boldsymbol{x}^*) + \lambda \nabla g(\boldsymbol{x}^*) = 0, \lambda \neq 0$$

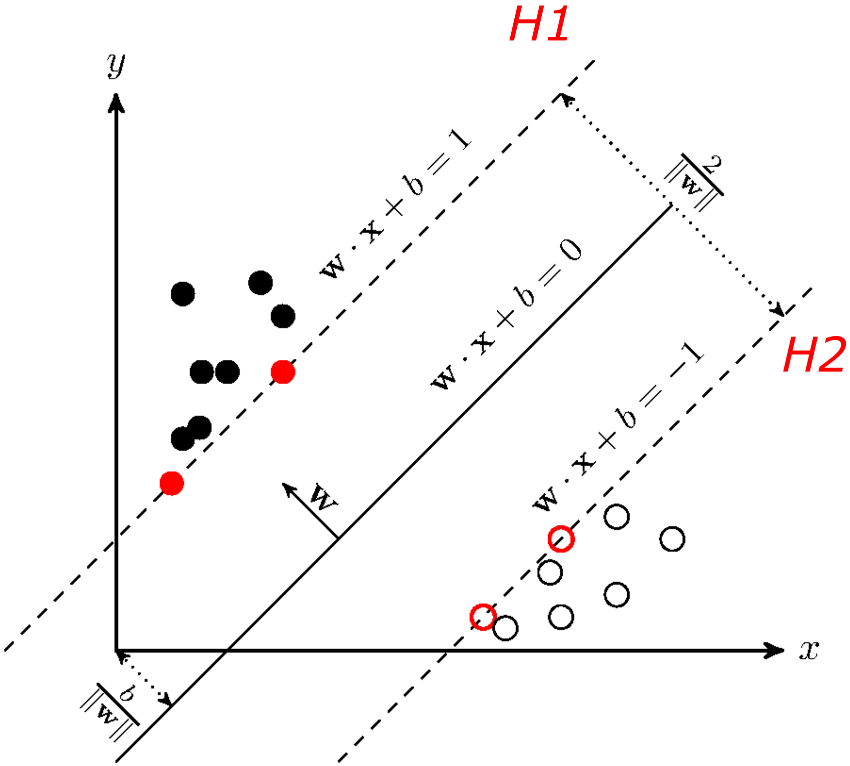
We can combine the constraints with objective function together

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$$

Margin

$$H1\text{:}\quad \boldsymbol{\omega}^{\boldsymbol{T}}\boldsymbol{x_i} + b = +1,$$

$$H2\text{:}\quad \boldsymbol{\omega}^{\boldsymbol{T}}\boldsymbol{x_i} + b = -1,$$



Recall that in 2-D, the distance from a point (x_0, y_0) to a line $Ax + By + C = 0$ is

$$\frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}$$

So the distance from hyperplane *H1* to H2 can be computed as

$$\frac{2}{||\boldsymbol{\omega}||}$$

Maximization

In conclusion, the objective function is

$$\max_{\boldsymbol{\omega},b} \frac{2}{||\boldsymbol{\omega}||}$$

$$\text{s. t.} \quad y_{\boldsymbol{i}}(\boldsymbol{\omega}^{\boldsymbol{T}}\boldsymbol{x}_{\boldsymbol{i}} + b) \geq 1, \qquad i=1,2,\ldots,\text{m}.$$

Alternatively, we can minimize the denominator

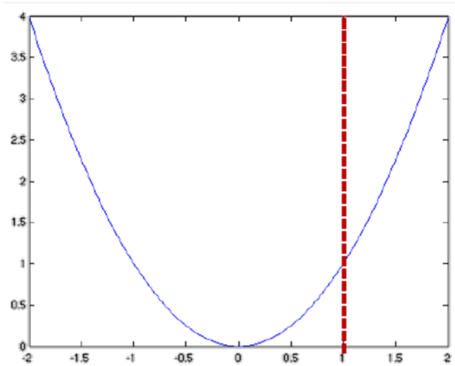
$$\min_{\boldsymbol{\omega},\boldsymbol{b}} \frac{1}{2}||\boldsymbol{\omega}||^2$$

$$\text{s. t.} \quad y_{\boldsymbol{i}}(\boldsymbol{\omega}^{\boldsymbol{T}}\boldsymbol{x}_{\boldsymbol{i}} + b) \geq 1, \qquad i=1,2,\ldots,\text{m}.$$

ML Estimation

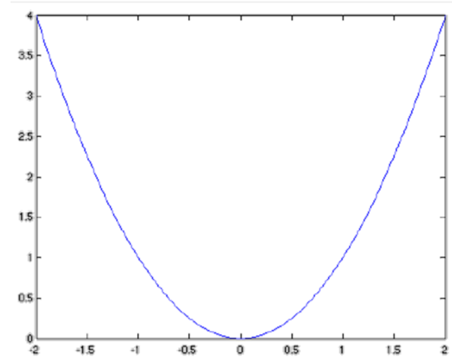
$$\min_x \; x^2$$
$$\text{s.t.} \quad x \geq b$$

$$\min_x \; x^2$$
$$\text{s.t.} \quad x \geq 1$$



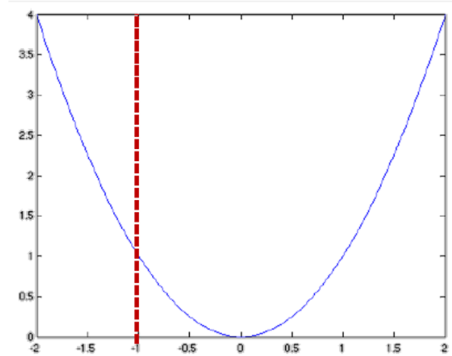
$$x^* = 1$$

$$\min_x \; x^2$$



$$x^* = 0$$

$$\min_x \; x^2$$
$$\text{s.t.} \quad x \geq -1$$



$$x^* = 0$$

Langrangian Muliplier

Move the constraint to objective function – **Lagrangian**

$$L(x, \alpha) = x^2 - \alpha(x - b), \quad \text{s.t.: } \alpha \geq 0$$

$$\begin{aligned} \min_x \max_\alpha \quad & L(x, \alpha) \\ \text{s.t.:} \quad & \alpha \geq 0 \end{aligned}$$

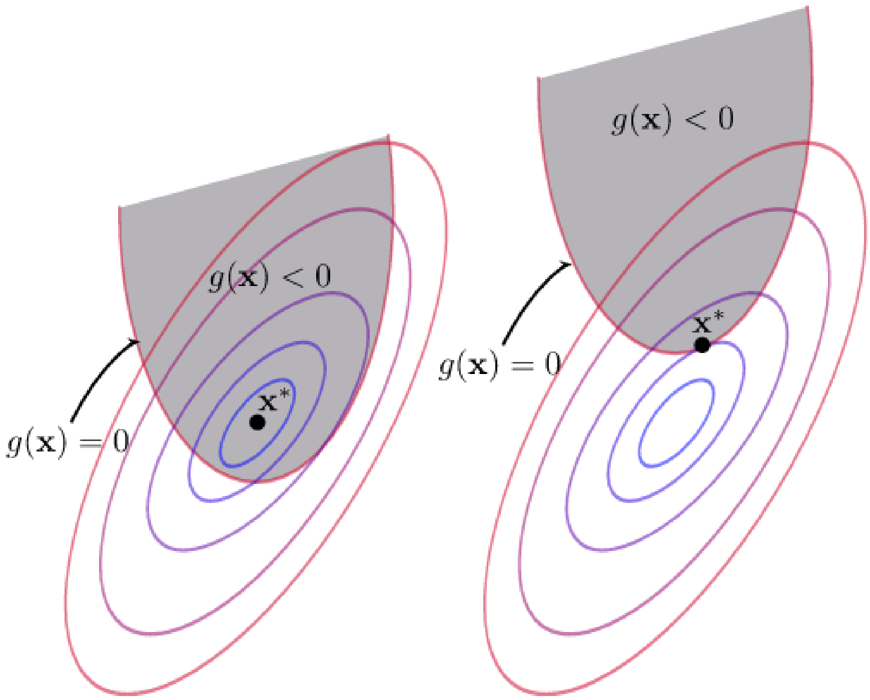
$$\begin{aligned} \min_x \max_\alpha \quad & L(x, \alpha) = x^2 - \alpha(x - b) \\ \text{s.t.:} \quad & \alpha \geq 0 \end{aligned}$$

To solve the min max problem

$$\frac{\partial L}{\partial x} = 0 \Rightarrow x^* = \frac{\alpha}{2}$$

$$\frac{\partial L}{\partial \alpha} = 0 \Rightarrow \alpha^* = \max(2b, 0)$$

Inequality Constraint



Case 2: inequality constraint

$$\min f(\boldsymbol{x})$$

$$s.\ b. \quad g(\boldsymbol{x}) \leq 0$$

$$\boxed{L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})}$$

$g(\boldsymbol{x}) < 0, \quad \lambda = 0$

$g(\boldsymbol{x}) = 0, \quad \nabla f(\boldsymbol{x}^*) + \lambda \nabla g(\boldsymbol{x}^*) = 0, \lambda > 0$

$\Rightarrow \lambda\, g(\boldsymbol{x}) = 0$

Dural Form

1. Primal problem

$$\min_{\boldsymbol{\omega},\boldsymbol{b}} \ \frac{1}{2}||\boldsymbol{\omega}||^2$$

$$\text{s. t.} \quad y_{\boldsymbol{i}}(\boldsymbol{\omega}^{\boldsymbol{T}}\boldsymbol{x}_{\boldsymbol{i}} + b) \geq 1, \qquad i\text{=1,2, …, m.}$$

2. Lagrange function

$$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2}||\boldsymbol{\omega}||^2 + \sum_{i=1}^{m} \alpha_i \left(1 - y_i\left(\boldsymbol{\omega}^T \boldsymbol{x}_i + b\right)\right)$$

$$\frac{\partial L(\boldsymbol{\omega},\ b,\ \boldsymbol{\alpha})}{\partial \boldsymbol{\omega}} = 0 \qquad \qquad \boldsymbol{\omega} = \sum_{i=1}^{m} \alpha_i y_i x_i$$

$$\Longrightarrow$$

$$\frac{\partial L(\boldsymbol{\omega},\ b,\ \boldsymbol{\alpha})}{\partial \mathrm{b}} = 0 \qquad \qquad 0 = \sum_{i=1}^{m} \alpha_i y_i$$

Dual Form

Move the constraint to objective function – **Lagrangian**

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x_i}^T \boldsymbol{x_j}$$

s.t. $\sum_{i=1}^{m} \alpha_i y_i = 0,$

$\alpha_i \geq 0, \qquad\qquad i = 1, \dots, m$

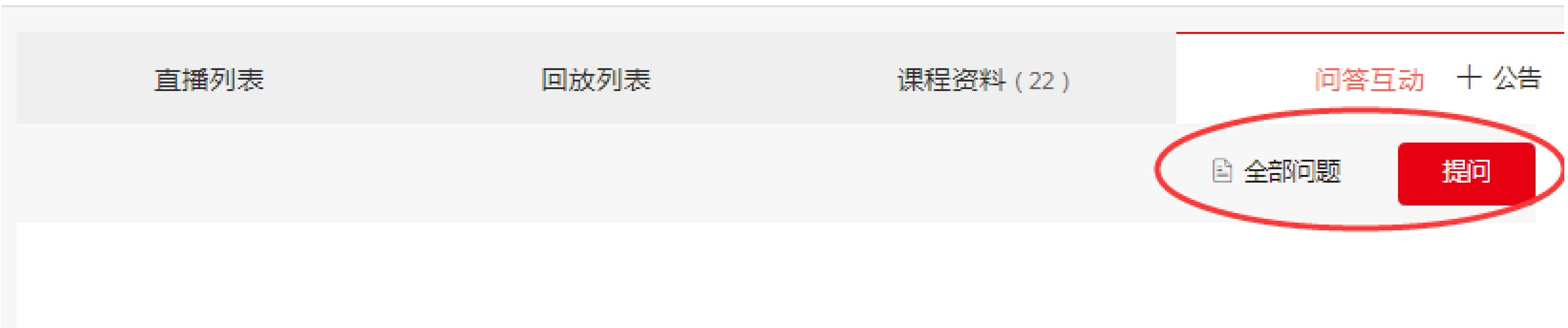
More details are available in the written lecture notes!

问答互动

PC端问答互动页面

1、点击“全部问题”显示本课程所有学员提问的问题。

2、点击“提问”即可向该课程的老师和助教提问问题。



联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：**小象学院**