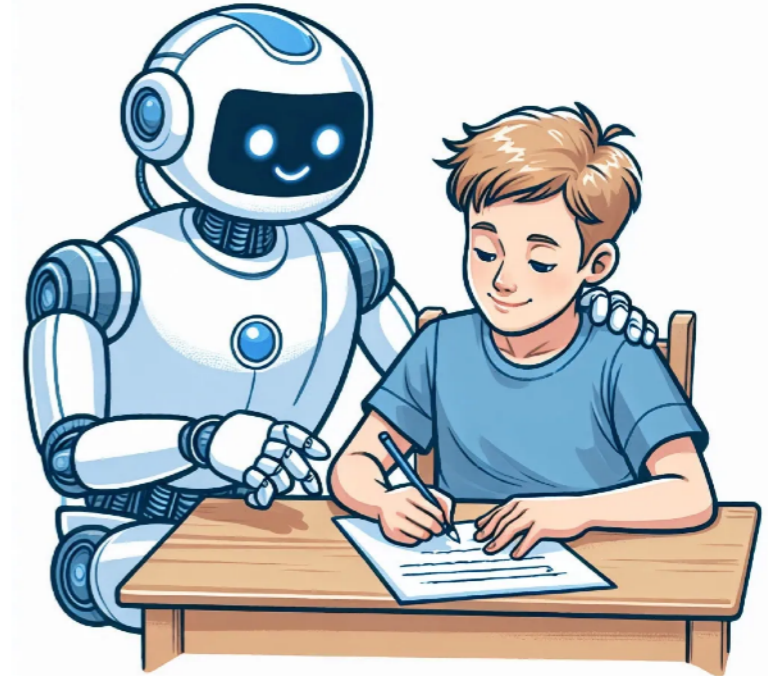


Content



0. Introduction

1. Regression

1.1 Multivariate Linear Regression (curve fitting)

1.2 Regularization (Lagrange multiplier)

1.3 Logistic Regression (Fermi-Dirac distribution)

1.4 Support Vector Machine (high-school geometry)

2. Dimensionality Reduction/feature extraction

2.1 Principal Component Analysis (order parameters)

2.2 Recommender Systems

2.3 Clustering (phase transition)

Content

3. Neural Networks

3.1 Biological neural networks

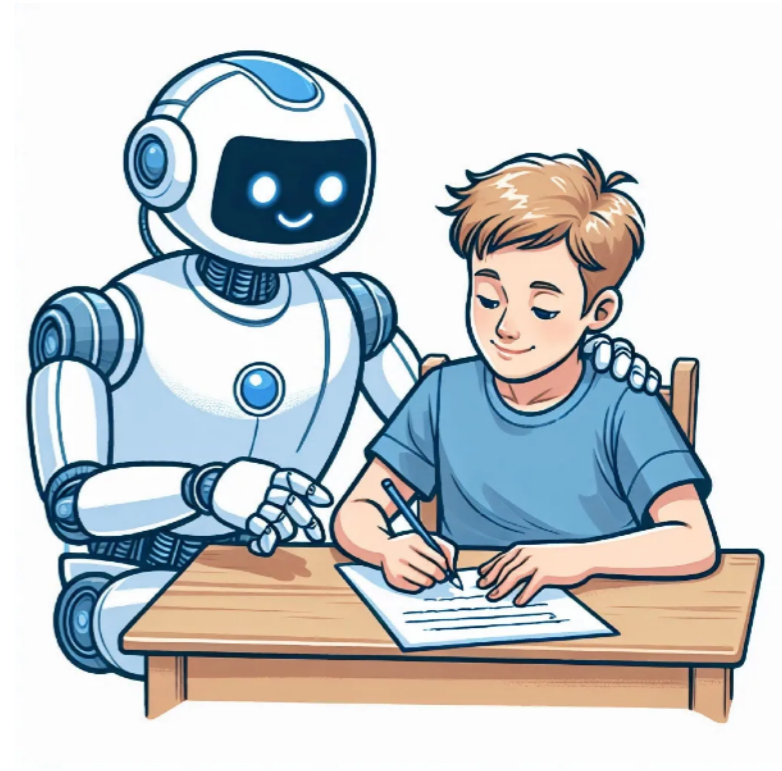
3.2 Mathematical representation

3.3 Factoring biological ingredient

3.4 Feed-forward neural networks

3.5 Learning algorithm

3.6 Universal Approximation Theorem

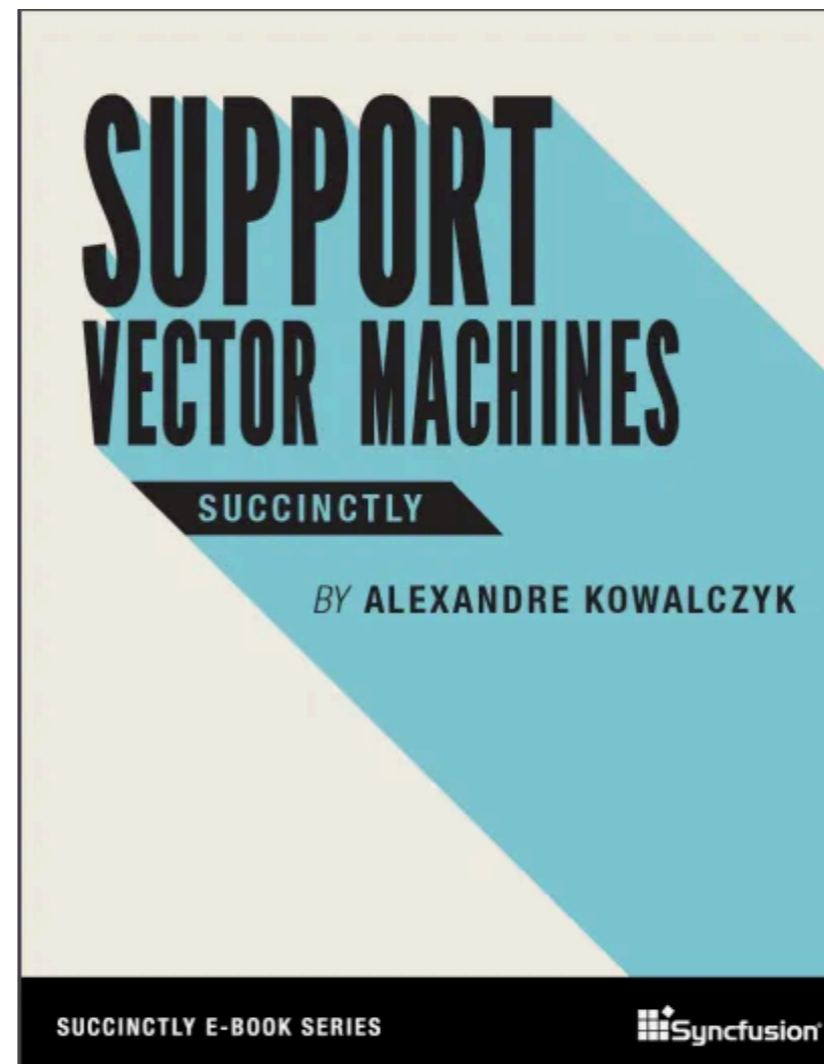


Support Vector Machine

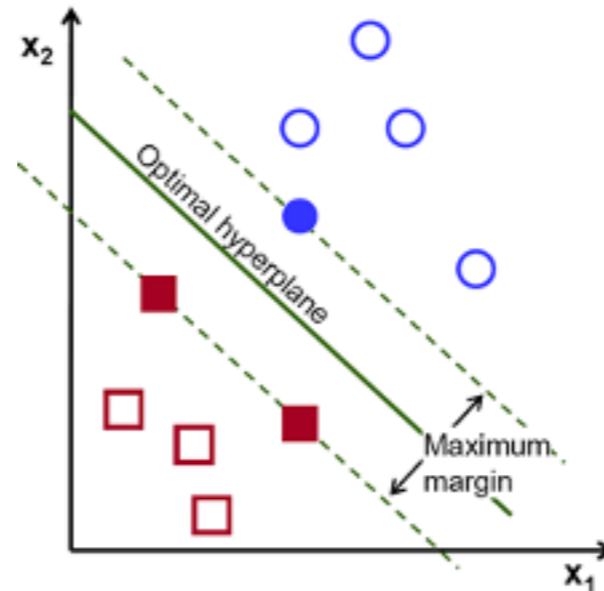
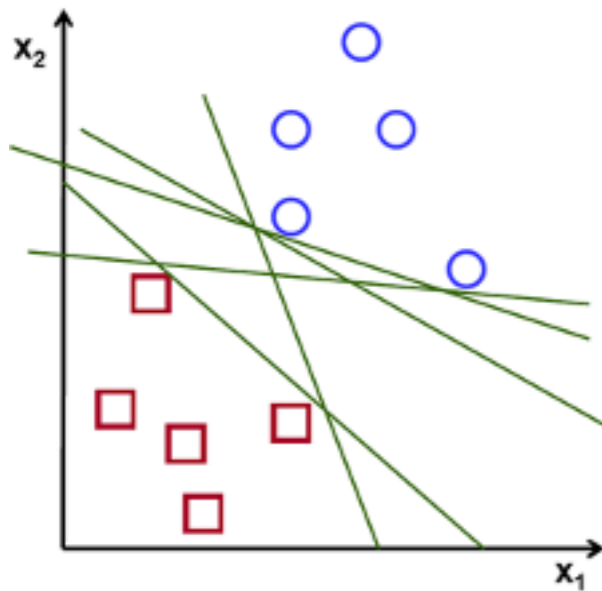
Good references by Alexandre Kowalczyk

<https://www.svm-tutorial.com/>

https://www.syncfusion.com/ebooks/support_vector_machines_succinctly

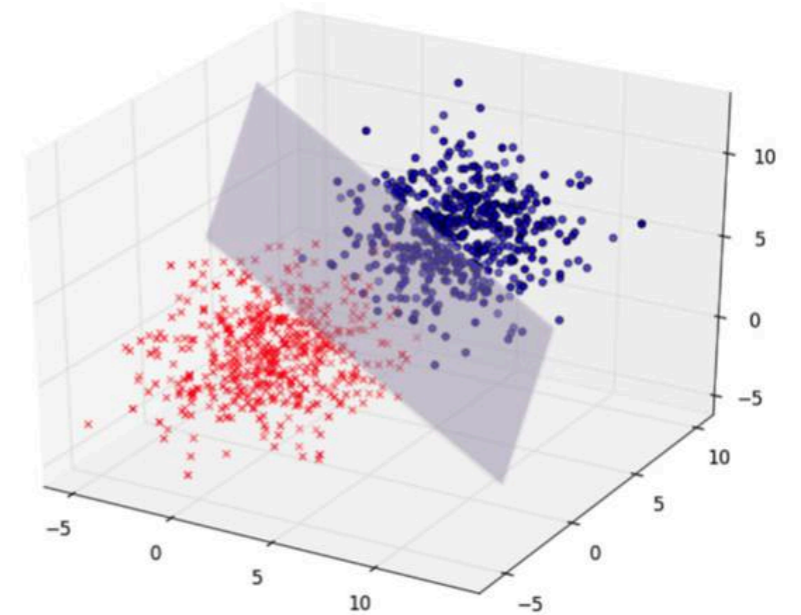
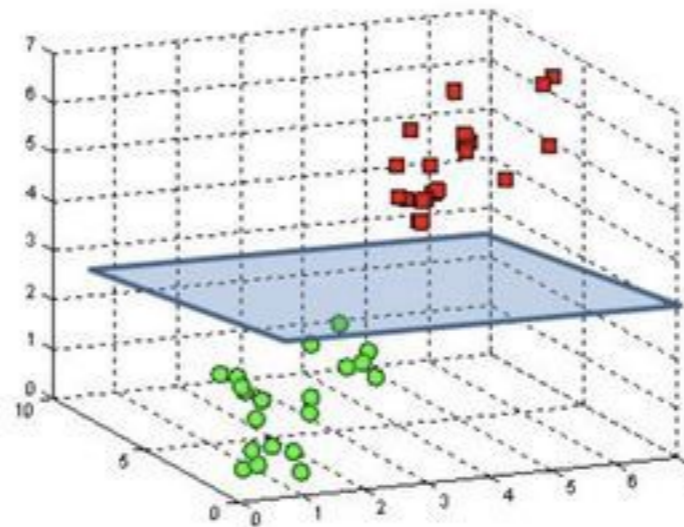
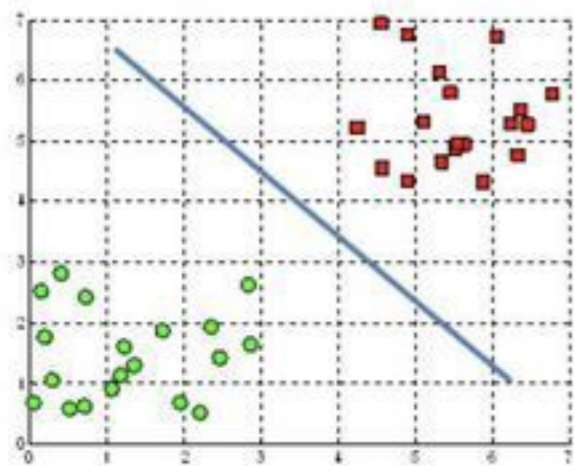


Support Vector Machine



A data point: p -dimensional vector ($p=2$)
Can be separated by $(p-1)$ -dimensional hyperplane

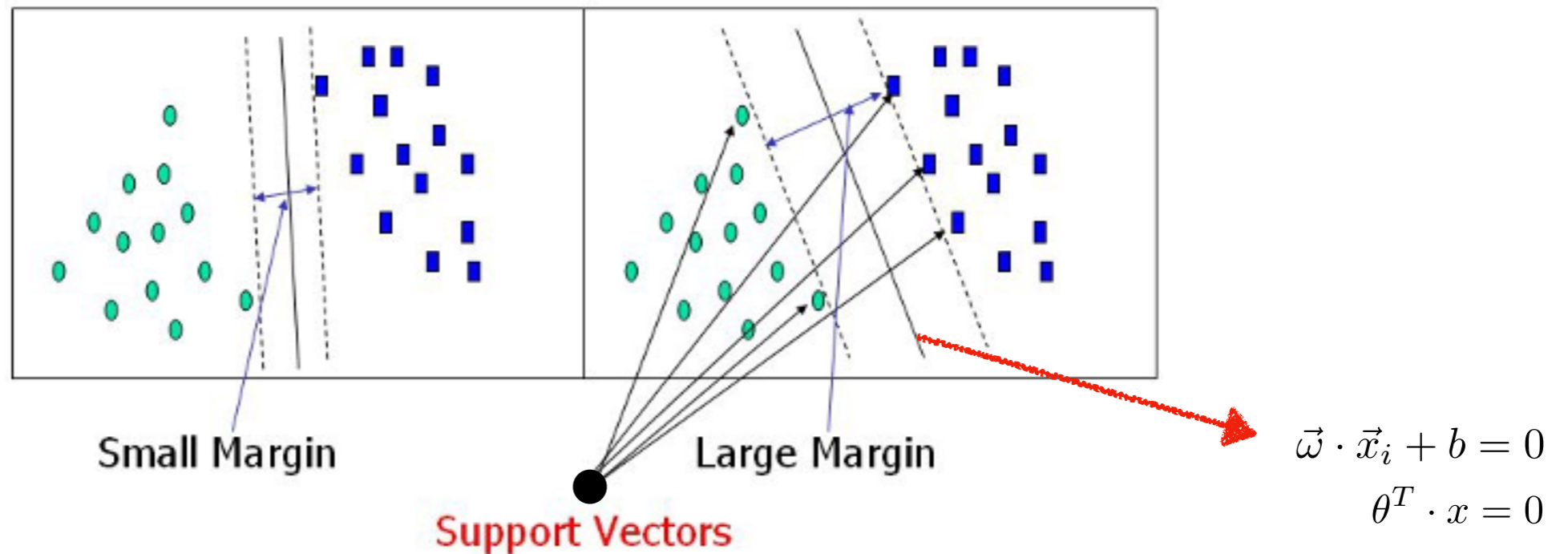
Hyperplane with maximal margin (largest separation between classes)



Hyperplane in R^2 is a line

Hyperplane in R^3 is a plane

Support Vector Machine



Optimisation problem with constraints:

Linearly separable training set $\mathcal{D} = \{(\vec{x}_i, y_i) | \vec{x}_i \in \mathbb{R}^n, y_i \in \{-1, 1\}\}_{i=1}^m$

Geometric margin $M = \min_{i=1,2,\dots,m} \frac{|y_i(\vec{\omega} \cdot \vec{x}_i + b)|}{\|\vec{\omega}\|}$

The optimal separating hyperplane is the hyperplane $(\vec{\omega}, b)$ whose margin M is the largest

Some high-school Geometry

Given a plane $w_1 x_1 + w_2 x_2 + b = 0$

* distance from origin $(0,0)$ to the plane is

$$\frac{b}{\sqrt{w_1^2 + w_2^2}} = \frac{b}{\|\vec{w}\|}$$

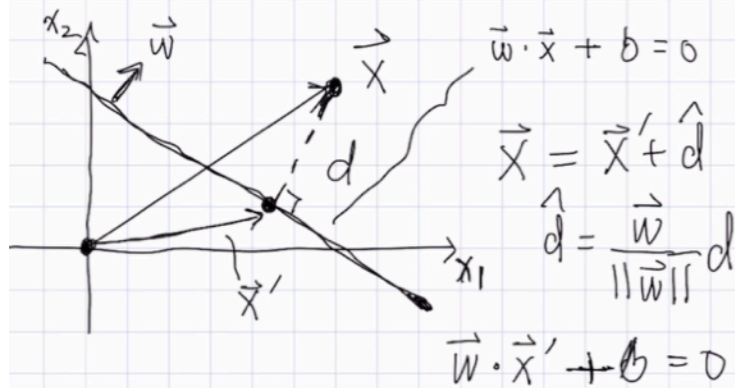
* distance from arbitrary point (x_1, x_2) to the plane is

$$\frac{|w_1 x_1 + w_2 x_2 + b|}{\|\vec{w}\|}$$

* $\vec{w} = (w_1, w_2)$ is the normal vector to the plane

$$\|\vec{w}\| = \sqrt{w_1^2 + w_2^2}$$

proof of distance from $\vec{x} = (x_1, x_2)$ to the plane $\vec{w} \cdot \vec{x} + b = 0$ $d = \frac{\vec{w} \cdot \vec{x} + b}{\|\vec{w}\|}$



$$\vec{w} \cdot (\vec{x} - \hat{d}) + b = 0$$

$$\vec{w} \cdot \vec{x} - \vec{w} \cdot \frac{\vec{w}}{\|\vec{w}\|} d + b = 0$$

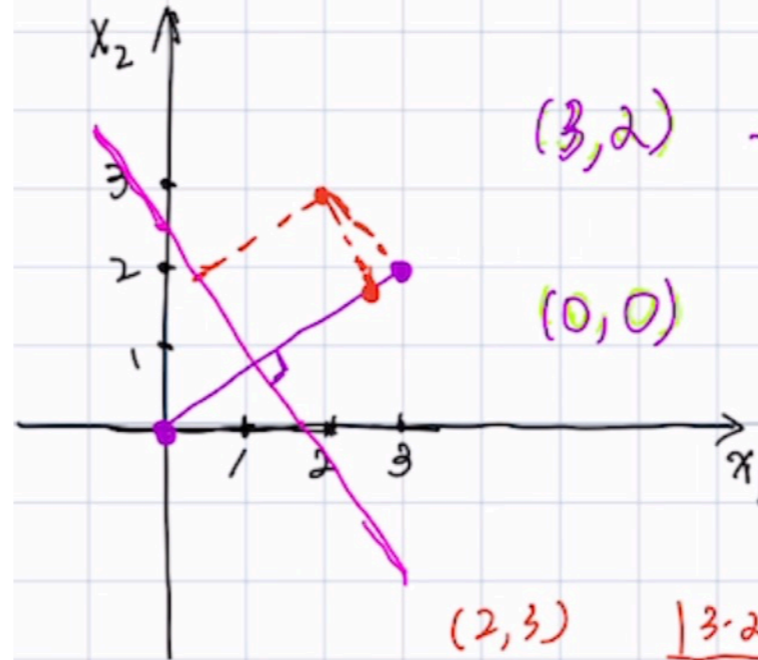
$$d = \frac{\vec{w} \cdot \vec{x} + b}{\|\vec{w}\|}$$

Example

$$3x_1 + 2x_2 - 5 = 0$$

normal vector $\vec{w} = (3, 2)$

$$\|\vec{w}\| = \sqrt{13}$$



$$(3, 2) \quad \frac{|3 \cdot 3 + 2 \cdot 2 - 5|}{\sqrt{13}} = \frac{8}{\sqrt{13}}$$

$$(0, 0) \quad \frac{|-5|}{\sqrt{13}} = \frac{5}{\sqrt{13}}$$

$$\frac{8}{\sqrt{13}} + \frac{5}{\sqrt{13}} = \sqrt{13} = \|\vec{w}\|$$

$$(2, 3) \quad \frac{|3 \cdot 2 + 2 \cdot 3 - 5|}{\sqrt{13}} = \frac{7}{\sqrt{13}}$$

distance between $(2, 3)$ and $(3, 2)$ $\sqrt{1+1} = \sqrt{2}$

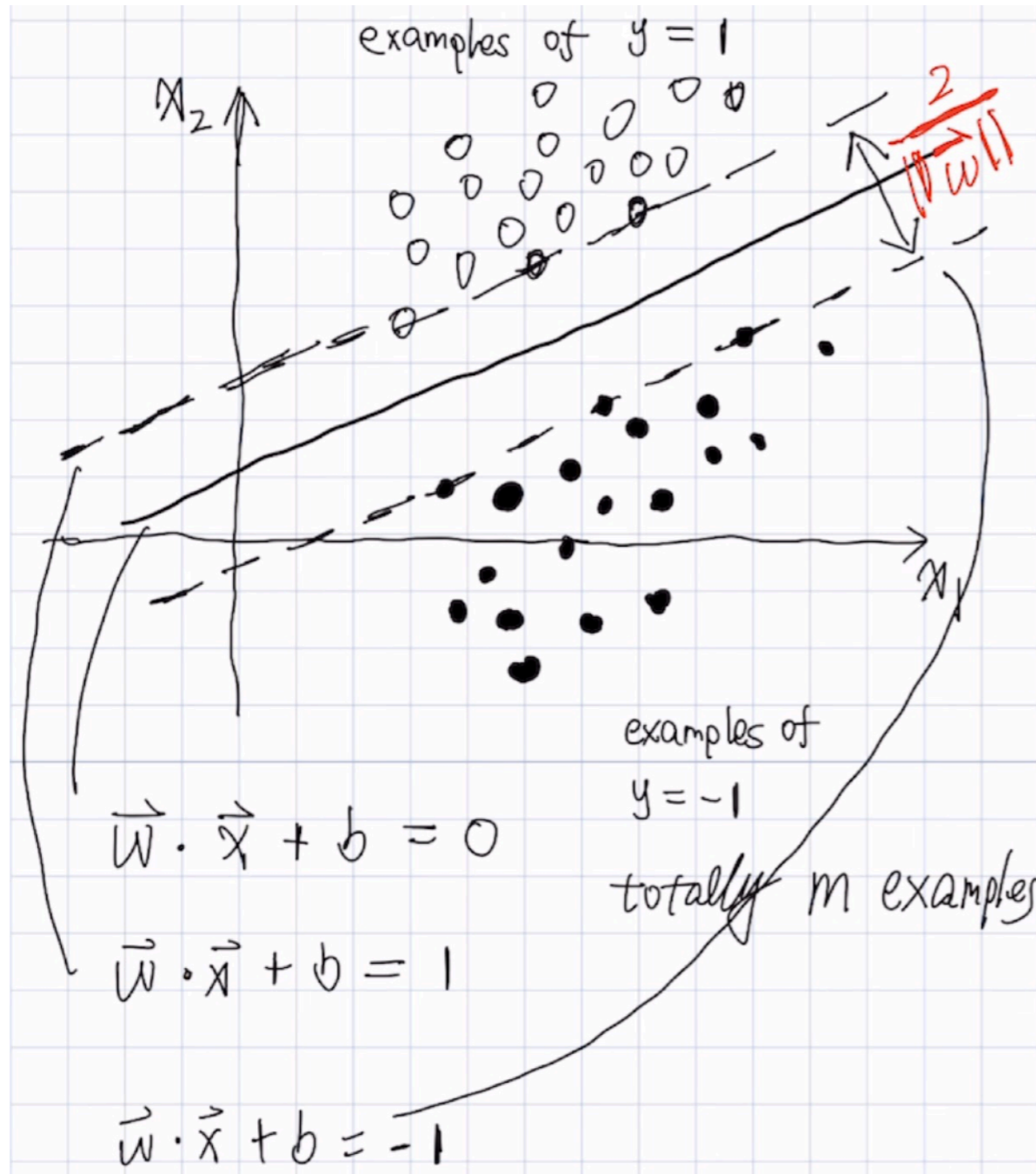
distance between $(2, 3)$ and $2x_1 - 3x_2 = 0$ is

$$\frac{|2 \cdot 2 - 3 \cdot 3|}{\sqrt{13}} = \frac{5}{\sqrt{13}}$$

$$\text{so } \sqrt{2} = \sqrt{\left(\frac{5}{\sqrt{13}}\right)^2 + \left(\frac{1}{\sqrt{13}}\right)^2} = \sqrt{2}$$

Support Vector Machine

How to find the optimal hyperplane for a dataset among all possible hyperplanes



Geometric margin M : distance/2 between two hyperplane

$$M = \frac{1}{\|\vec{w}\|}$$

Maximize the geometric margin means minimize

$$\|\vec{w}\|$$

Constraints: at the same time, prevent data points from falling into the margin

To find (\vec{w}, b) such that

$$\begin{aligned} & \max_{(\vec{w}, b)} M \\ & \text{subject to } \frac{|y_i(\vec{w} \cdot \vec{x}_i + b)|}{\|\vec{w}\|} \geq M, \quad i = 1, 2, \dots, m \end{aligned}$$

Constrained optimization problem

Support Vector Machine

To find $(\vec{\omega}, b)$ such that

$$\begin{aligned} & \max_{(\vec{\omega}, b)} M \\ \text{subject to } & \frac{|y_i(\vec{\omega} \cdot \vec{x}_i + b)|}{\|\vec{\omega}\|} \geq M, \quad i = 1, 2, \dots, m \end{aligned}$$

Is equivalent to the minimisation problem with constraints, remember $M = \frac{1}{\|\vec{\omega}\|}$

$$\begin{aligned} & \min_{(\vec{\omega}, b)} \|\vec{\omega}\| \\ \text{subject to } & y_i(\vec{\omega} \cdot \vec{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m \end{aligned}$$

Is equivalent to

$$\begin{aligned} & \min_{(\vec{\omega}, b)} \frac{1}{2} \|\vec{\omega}\|^2 \\ \text{subject to } & y_i(\vec{\omega} \cdot \vec{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m \end{aligned}$$

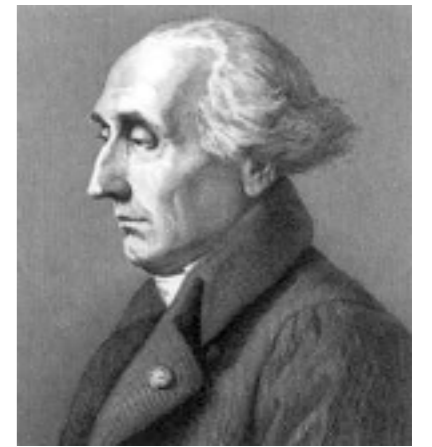


Lagrange multipliers and duality



Convex quadratic optimisation problem

Lagrange multipliers and duality



Joseph-Louis Lagrange (1736-1813)


$$\begin{array}{l} \text{minimize } f(\vec{x}) \\ \text{subject to } g(\vec{x}) = 0 \\ \text{equality constraint} \end{array}$$

find the local maxima and minima

$$\mathcal{L} = f(\vec{x}) - \alpha g(\vec{x})$$


$$\nabla \mathcal{L}(\vec{x}, \alpha) = \nabla f(\vec{x}) - \alpha \nabla g(\vec{x}) = 0$$

$$\begin{array}{l} \text{minimize}_{x,y} \quad f(x,y) = x^2 + y^2 \\ \text{subject to} \quad g(x,y) = x + y - 1 = 0 \end{array}$$



“I will deduce the complete mechanics of solid and fluid bodies using the principle of least action.”

JOSEPH-LOUIS LAGRANGE
Letter to Leonhard Euler, May 1756



“I have almost completed a book on analytical mechanics founded solely on the principle [of virtual work]. But since I still have no idea where and when it can be published, I am not in any hurry to finish it.”

JOSEPH-LOUIS LAGRANGE
Letter to Pierre Laplace, September 1782

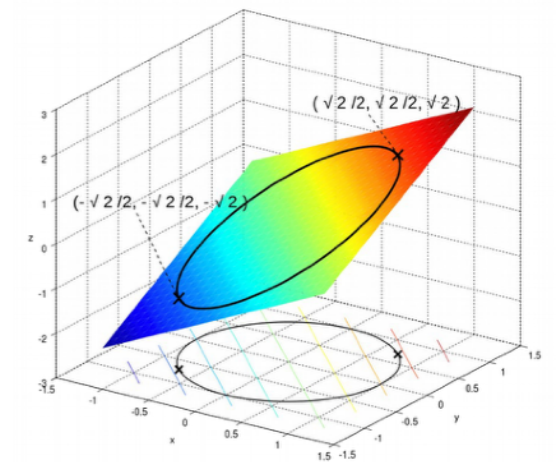
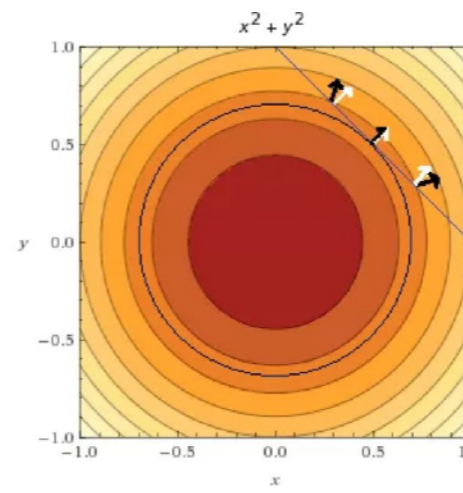
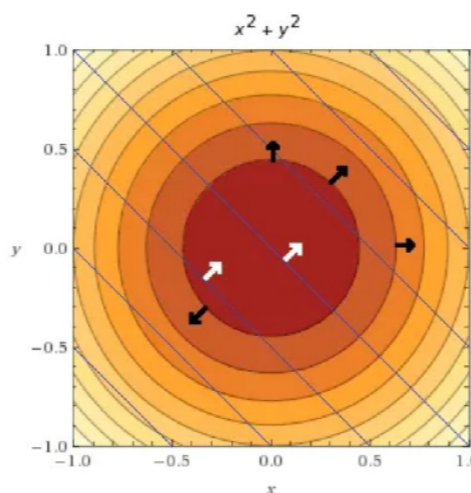
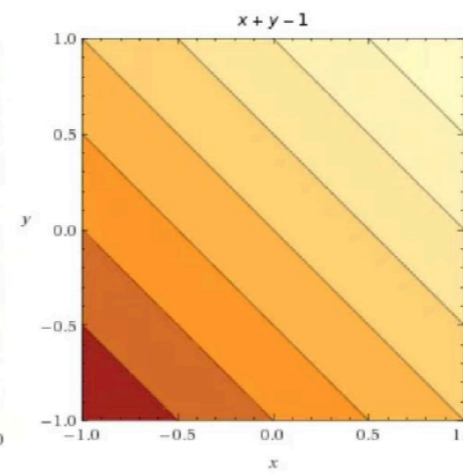
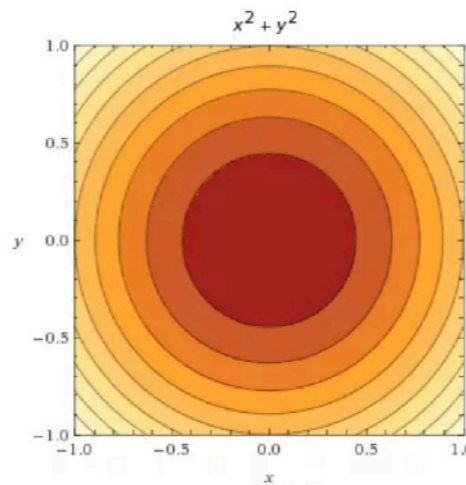
$$\mathcal{L}(x, y, \alpha) = f(x, y) - \alpha g(x, y)$$

$$\nabla \mathcal{L}(x, y, \alpha) = \nabla f(x, y) - \alpha \nabla g(x, y) = 0$$

$$\nabla_{x_1, x_2, \dots, x_n, \alpha} \mathcal{L}(x_1, x_2, \dots, x_n, \alpha) = 0$$

$$\frac{\partial \mathcal{L}}{\partial x} = 0 \quad \frac{\partial \mathcal{L}}{\partial y} = 0 \quad \frac{\partial \mathcal{L}}{\partial \alpha} = 0$$

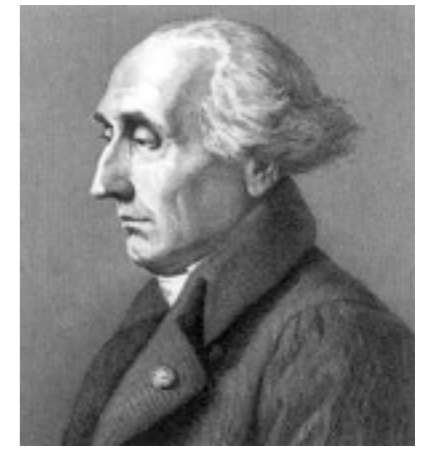
$$x = y = \frac{1}{2} \quad \alpha = 1$$



$$f(x, y) = x + y$$

$$g(x, y) = x^2 + y^2 - 1 = 0$$

Lagrange multipliers and duality



Joseph-Louis Lagrange (1736-1813)

$$\begin{array}{l} \text{minimize } f(\vec{x}) \\ \text{subject to } g(\vec{x}) = 0 \\ \text{equality constraint} \end{array}$$

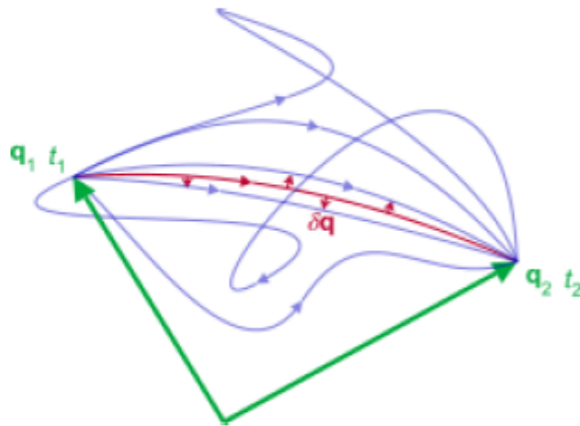
$$\mathcal{L} = f(\vec{x}) - \alpha g(\vec{x})$$

$$\nabla \mathcal{L}(\vec{x}, \alpha) = \nabla f(\vec{x}) - \alpha \nabla g(\vec{x}) = 0$$

$$L = T - V$$

Euler-Lagrange equation


$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_j} \right) = \frac{\partial L}{\partial q_j}$$



As the system evolves, \mathbf{q} traces a path through configuration space (only some are shown). The path taken by the system (red) has a stationary action ($\delta S = 0$) under small changes in the configuration of the system ($\delta \mathbf{q}$).^[27]


$$S = \int_{t_1}^{t_2} L dt, \quad \delta S = 0$$

principle of least action.



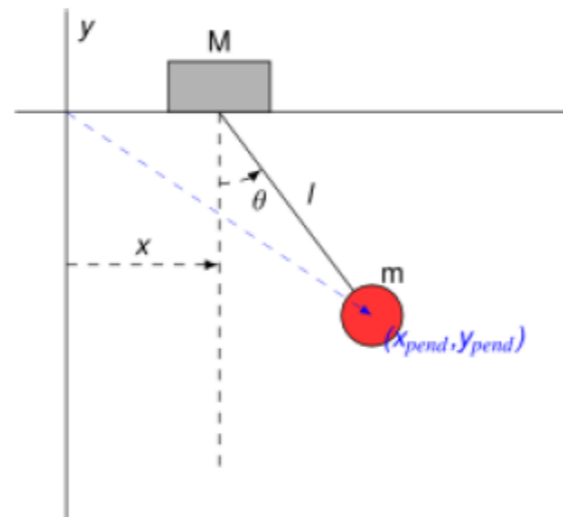
“I will deduce the complete mechanics of solid and fluid bodies using the principle of least action.”

JOSEPH-LOUIS LAGRANGE
Letter to Leonhard Euler, May 1756



“I have almost completed a book on analytical mechanics founded solely on the principle [of virtual work]. But since I still have no idea where and when it can be published, I am not in any hurry to finish it.”

JOSEPH-LOUIS LAGRANGE
Letter to Pierre Laplace, September 1782



Sketch of the situation with definition of the coordinates (click to enlarge)

$$T = \frac{1}{2} M \dot{x}^2 + \frac{1}{2} m (\dot{x}_{\text{pend}}^2 + \dot{y}_{\text{pend}}^2)$$

$$V = mgy_{\text{pend}}$$

$$L = T - V$$

$$= \frac{1}{2} M \dot{x}^2 + \frac{1}{2} m \left[(\dot{x} + l\dot{\theta} \cos \theta)^2 + (l\dot{\theta} \sin \theta)^2 \right] + mgl \cos \theta$$

$$\frac{d}{dt} \left[m(\dot{x}l \cos \theta + l^2 \dot{\theta}) \right] + ml(\dot{x}\dot{\theta} + g) \sin \theta = 0$$

$$\ddot{\theta} + \frac{\ddot{x}}{l} \cos \theta + \frac{g}{l} \sin \theta = 0$$

Lagrange multipliers and duality

$$\min_{(\vec{\omega}, b)} \frac{1}{2} \|\vec{\omega}\|^2$$

subject to $y_i(\vec{\omega} \cdot \vec{x}_i + b) \geq 1, i = 1, 2, \dots, m$

$$\mathcal{L}(\vec{\omega}, b, \alpha) = \frac{1}{2} \|\vec{\omega}\|^2 - \sum_{i=1}^m \alpha_i [y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1]$$

$$\min_{(\vec{\omega}, b)} \max_{\alpha} \mathcal{L}(\vec{\omega}, b, \alpha)$$

subject to $\alpha_i \geq 0, i = 1, 2, \dots, m$

$$\nabla_{\vec{\omega}} \mathcal{L} = \vec{\omega} - \sum_{i=1}^m \alpha_i y_i \vec{x}_i = 0$$

$$\frac{\partial \mathcal{L}}{\partial b} = - \sum_{i=1}^m \alpha_i y_i = 0$$

$$\mathcal{L}_D = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j$$

$$\max_{\alpha} \mathcal{L}_D(\alpha, \vec{x}_i, y_i)$$

subject to $\alpha_i \geq 0, i = 1, 2, \dots, m$

$$\sum_{i=1}^m \alpha_i y_i = 0$$

gradient, not the -gradient

Karush-Kuhn-Tucker (KKT) conditions

- Stationarity condition:

$$\nabla_{\vec{\omega}} \mathcal{L} = \vec{\omega} - \sum_{i=1}^m \alpha_i y_i \vec{x}_i = 0$$

$$\frac{\partial \mathcal{L}}{\partial b} = - \sum_{i=1}^m \alpha_i y_i = 0$$

- Primal feasibility condition:

$$y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1 \geq 0 \quad \text{for all } i = 1, 2, \dots, m$$

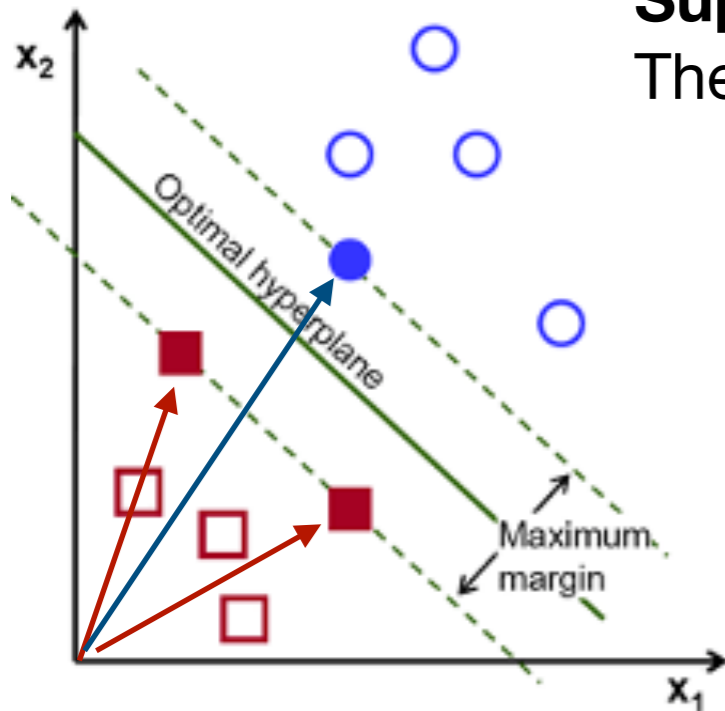
- Dual feasibility condition:

$$\alpha_i \geq 0 \quad \text{for all } i = 1, 2, \dots, m$$

- Complementary slackness condition:

$$\alpha_i [y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1] = 0 \quad \text{for all } i = 1, 2, \dots, m$$

Support vectors are examples having a positive Lagrange multiplier. They are the ones the constraint is active.



Once have the multipliers and support vectors

$$\vec{\omega} = \sum_{i=1}^m \alpha_i y_i \vec{x}_i$$

$$b = \frac{1}{S} \sum_{i=1}^S (y_i - \vec{\omega} \cdot \vec{x}_i)$$

S is the number of support vectors

Support Vector Machine: Hinge Loss

$$\mathcal{L}(\vec{\omega}, b, \alpha) = \frac{1}{2} \|\vec{\omega}\|^2 - \sum_{i=1}^m \alpha_i [y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1]$$

$$J(\vec{\omega}, b) = \frac{1}{2m} \|\vec{\omega}\|^2 + \frac{1}{m} \sum_{i=1}^m \max(0, 1 - y_i(\vec{\omega} \cdot \vec{x}_i + b))$$

Replace $\frac{1}{2m}$ with $\frac{\lambda}{2m}$ for regularisation

$$y_i = 1 \quad \vec{\omega} \cdot \vec{x}_i + b > 1 \quad \nabla_{\vec{\omega}} J = \frac{1}{m} \vec{\omega}$$

$$\vec{\omega} \cdot \vec{x}_i + b < 1 \quad \nabla_{\vec{\omega}} J = \frac{1}{m} \vec{\omega} - y_i \vec{x}_i$$

$$\nabla_b J = -y_i$$

Gradient Descent

$$\vec{\omega} = \vec{\omega} - \alpha \nabla_{\vec{\omega}} J$$

$$b = b - \alpha \nabla_b J$$

$$y_i = -1 \quad \vec{\omega} \cdot \vec{x}_i + b < -1 \quad \nabla_{\vec{\omega}} J = \frac{1}{m} \vec{\omega}$$

$$\vec{\omega} \cdot \vec{x}_i + b > -1 \quad \nabla_{\vec{\omega}} J = \frac{1}{m} \vec{\omega} - y_i \vec{x}_i$$

$$\nabla_b J = -y_i$$

Support Vector Machine: Hinge Loss

$$J(\theta) = -\frac{1}{M} \sum_{i=1}^M [y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)}))]$$

$$J(\theta) = \frac{1}{M} \sum_{i=1}^M \max(0, 1 - y_i(\vec{\theta} \cdot \vec{x}_i + \theta_0))$$

Hinge function

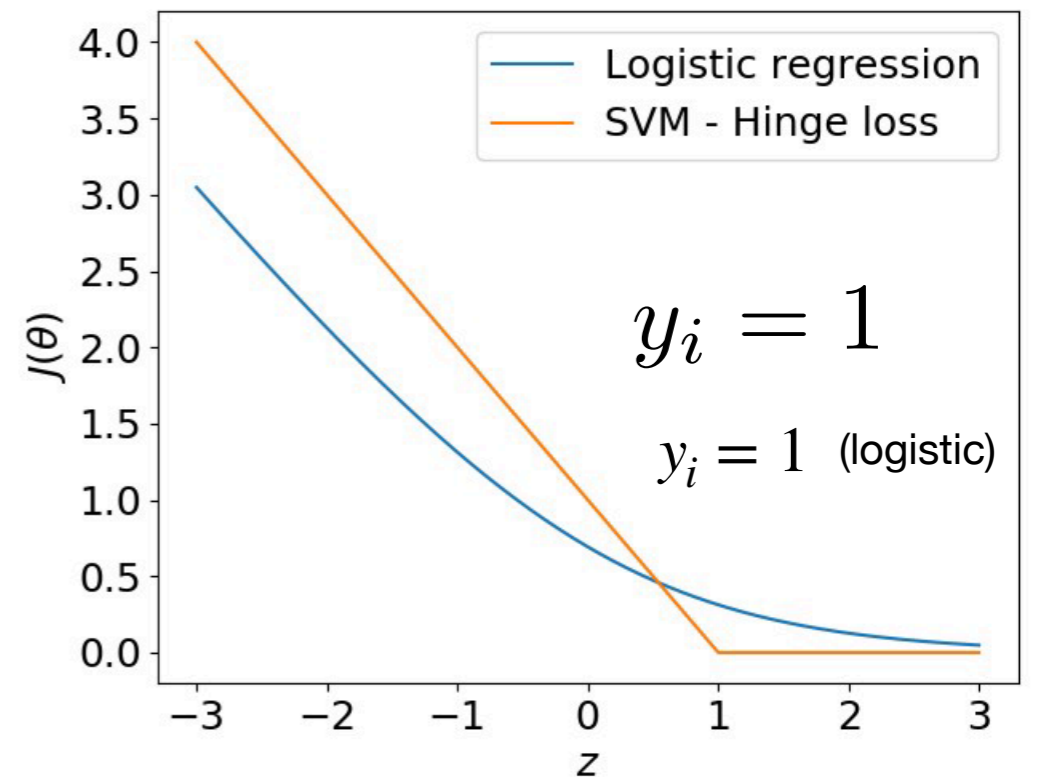
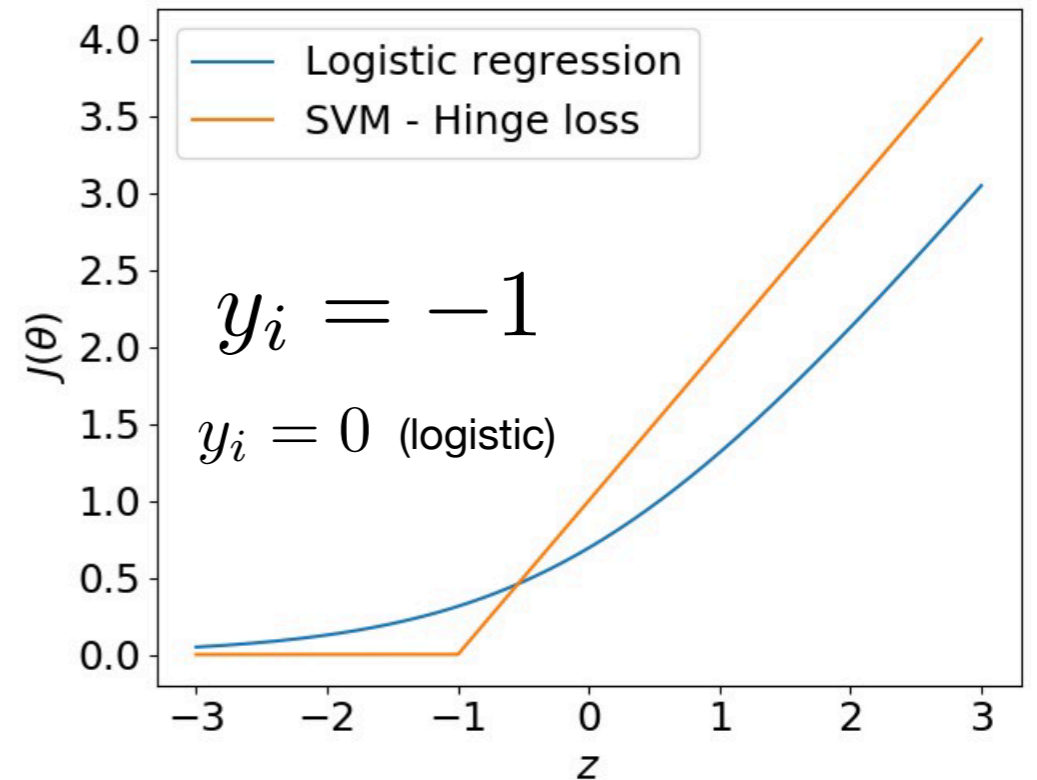
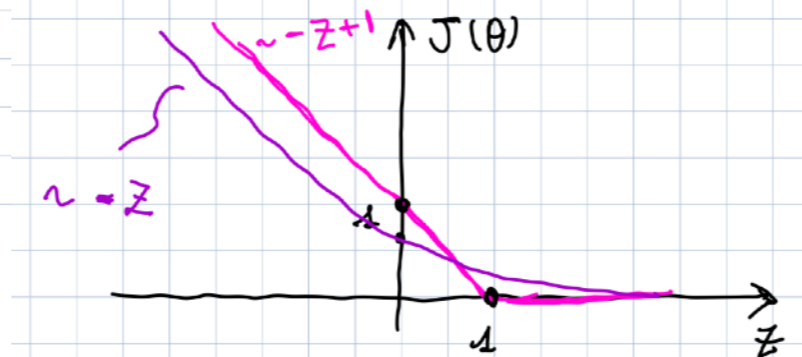
$$y_i = 1 \quad \vec{\theta} \cdot \vec{x}_i + \theta_0 \geq 1 \quad J(\theta) = 0$$

$$\vec{\theta} \cdot \vec{x}_i + \theta_0 < 1 \quad J(\theta) = 1 - (\vec{\theta} \cdot \vec{x}_i + \theta_0)$$

remember in logistic regression

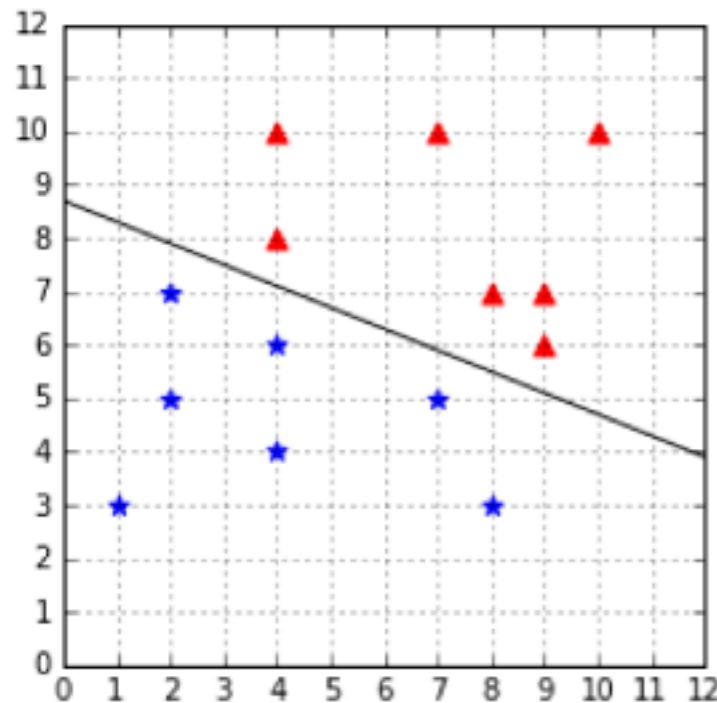
$$J(\theta) = -\log \frac{1}{1 + e^{-\theta \cdot x}}$$

$$z = \theta \cdot x = \vec{\theta} \cdot \vec{x} + \theta_0$$

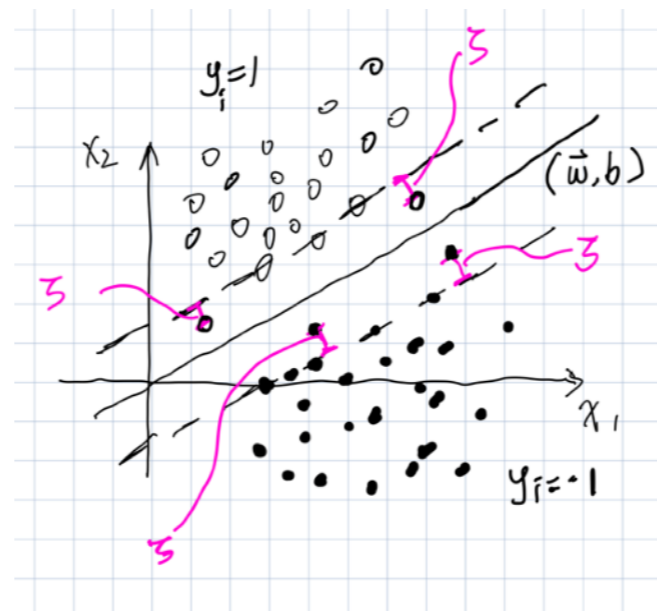
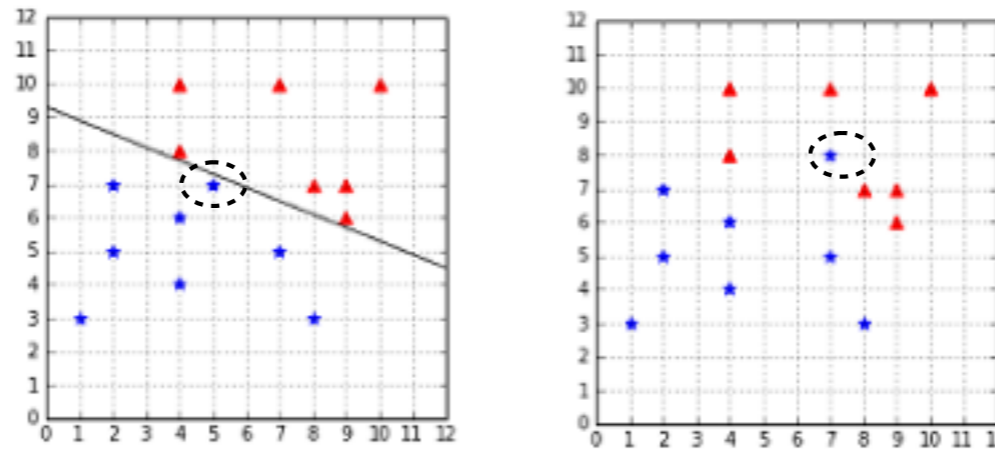


Support Vector Machine: hard margin and soft margin

Linearly separable



Outliers



Soft margin to rescue

$$y_i(\vec{\omega} \cdot \vec{x}_i + b) \geq 1$$

$$y_i(\vec{x} \cdot \vec{x}_i + b) \geq 1 - \zeta_i$$

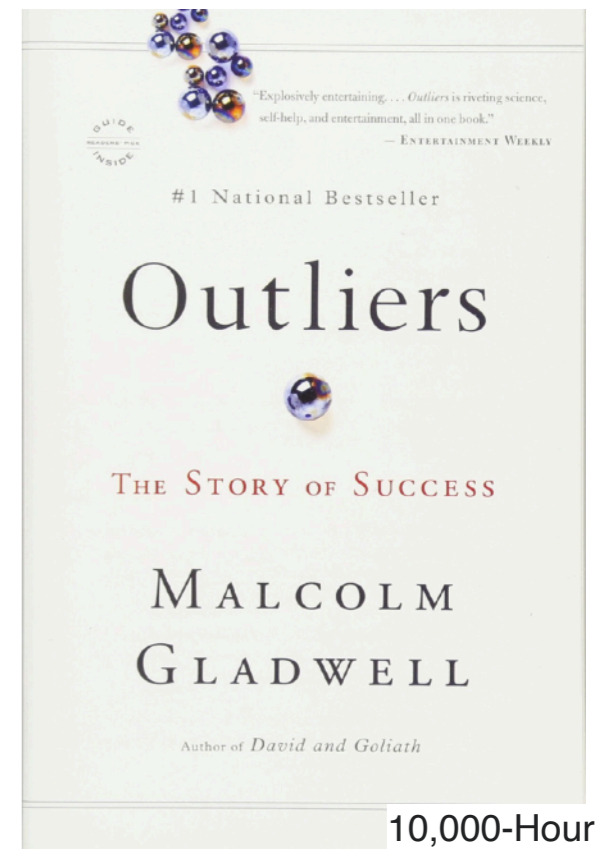
slack variables

Modify the objective function with regularization

$$\text{minimize}_{(\vec{\omega}, b, \zeta)} \quad \frac{1}{2} \|\vec{\omega}\|^2 + C \sum_{i=1}^m \zeta_i$$

$$\text{subject to } y_i(\vec{\omega} \cdot \vec{x}_i + b) \geq 1 - \zeta_i$$

$$\zeta_i \geq 0 \text{ for any } i = 1, \dots, m$$



10,000-Hour Rule

Lagrange multipliers and duality

$$\mathcal{L}(\vec{\omega}, b, \alpha, \zeta) = \frac{1}{2} \|\vec{\omega}\|^2 + C \sum_{i=1}^m \zeta_i - \sum_{i=1}^m \alpha_i [y_i (\vec{\omega}_i \cdot \vec{x}_i + b) - 1 + \zeta_i]$$

$$\nabla_{\vec{\omega}} \mathcal{L} = \vec{\omega} - \sum_{i=1}^m \alpha_i y_i \vec{x}_i = 0$$

Karush-Kuhn-Tucker (KKT) conditions

$$\frac{\partial \mathcal{L}}{\partial b} = - \sum_{i=1}^m \alpha_i y_i = 0$$

$$\frac{\partial \mathcal{L}}{\partial \zeta} = C - \alpha = 0$$

$$\mathcal{L}_D = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j$$

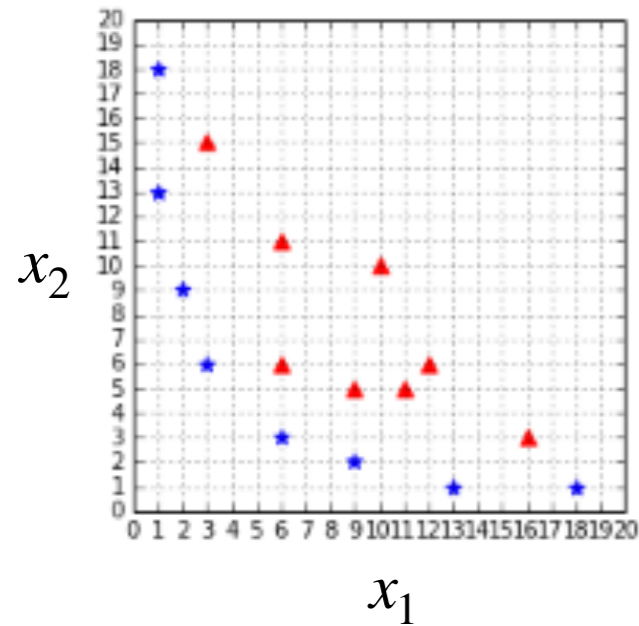
$$\max_{\alpha} \mathcal{L}_D(\alpha, \vec{x}_i, y_i)$$

subject to $0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, m$

$$\sum_{i=1}^m \alpha_i y_i = 0$$

Kernel machine: Dimensionality reduction strike

Not linearly separable in **two dimensions**

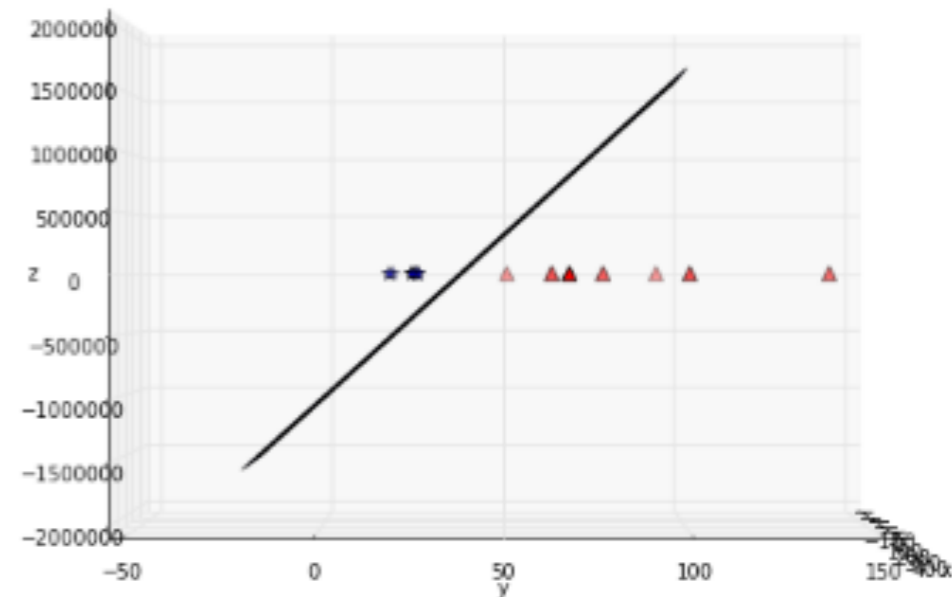
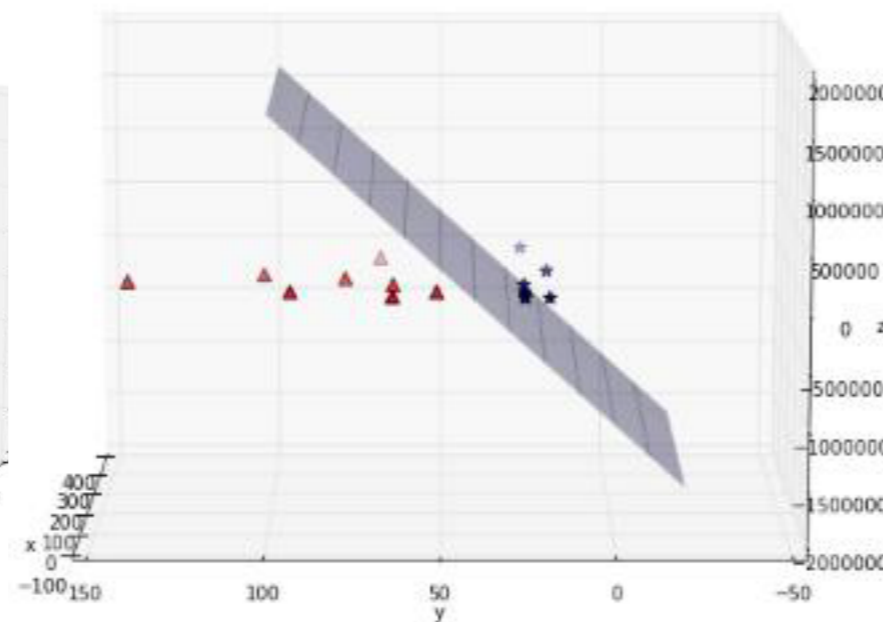
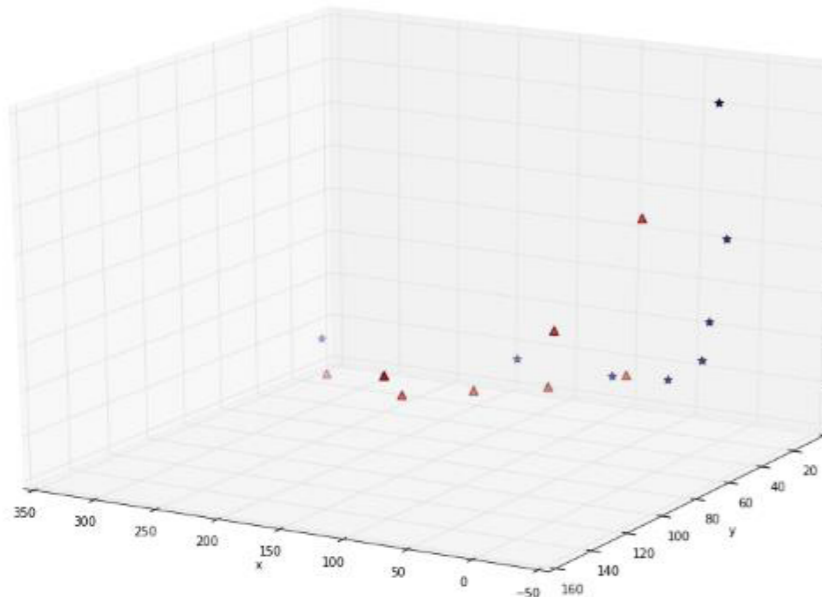


Polynomial mapping $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \quad \phi(x_1, x_2) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$

$$\langle \phi(\vec{x}_i), \phi(\vec{x}_j) \rangle_{\mathbb{R}^3} = (x_{i,1}^2, \sqrt{2}x_{i,1}x_{i,2}, x_{i,2}^2) \cdot (x_{j,1}^2, \sqrt{2}x_{j,1}x_{j,2}, x_{j,2}^2)$$

$$= x_{i,1}^2x_{i,2}^2 + 2x_{i,1}x_{i,2}x_{j,1}x_{j,2} + x_{i,2}^2x_{j,2}^2$$

$$= (\vec{x}_i \cdot \vec{x}_j + c)^d \quad \text{with } c = 0 \text{ and } d = 2$$



Kernel machine: Dimensionality reduction strike

mapping $\phi : \mathcal{X} \rightarrow \mathcal{V}$

function $K : \mathcal{X} \rightarrow \mathbb{R} \quad K(\vec{x}, \vec{x}') = \langle \phi(\vec{x}), \phi(\vec{x}') \rangle_{\mathcal{V}}$

inner production in \mathcal{V} , **kernel function**

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j K(\vec{x}_i, \vec{x}_j)$$

subject to $0 \leq \alpha_i \leq C, i = 1, \dots, m$

$$\sum_{i=1}^m \alpha_i y_i = 0$$

$$K(\vec{x}, \vec{x}') = (\vec{x} \cdot \vec{x}' + c)^d$$

$c = 1, d = 1$ linear kernel

$c = 0, d = 2$ quadratic kernel

