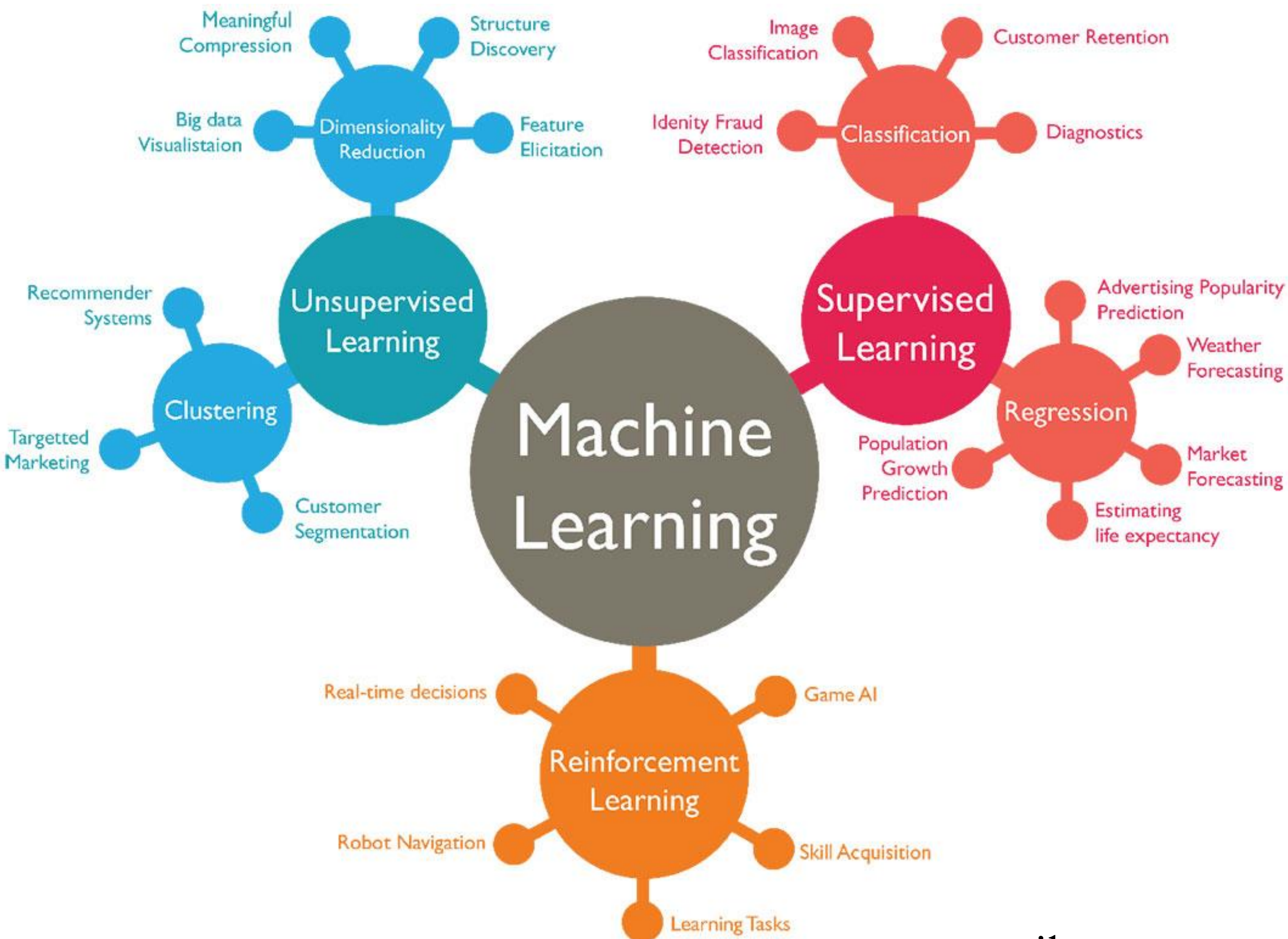
Trường Đại học Tôn Đức Thắng

Machine Learning
Support Vector Machine

Giảng viên: Tiến sĩ Bùi Thanh Hùng
Trưởng Lab Khoa học Phân tích dữ liệu và Trí tuệ nhân tạo
Giám đốc chương trình Hệ thống thông tin
Đại học Thủ Dầu Một
Email: tuhungphe@gmail.com
Website: https://sites.google.com/site/hungthanhbui1980/

3- Machine Learning

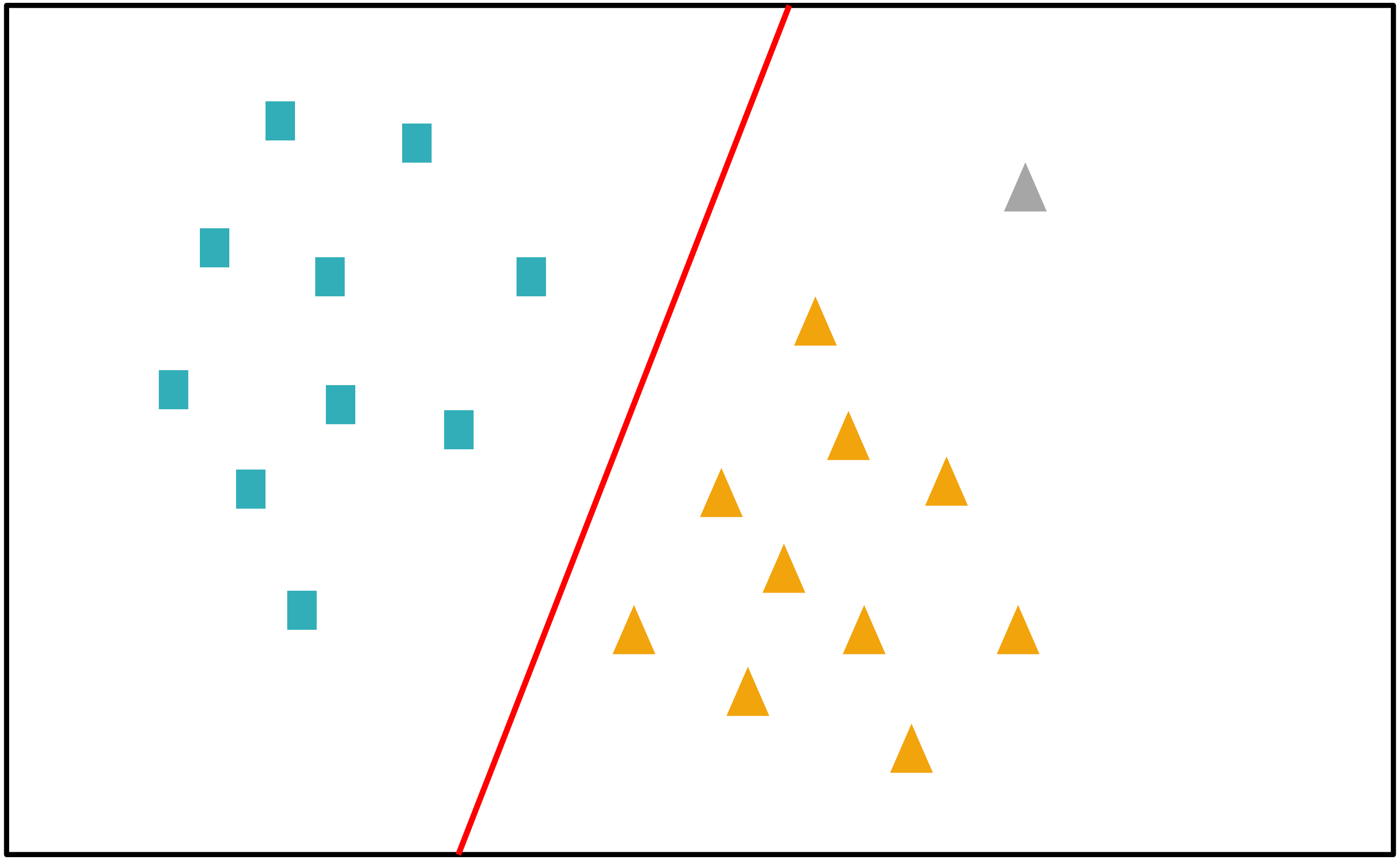


Support Vector Machine

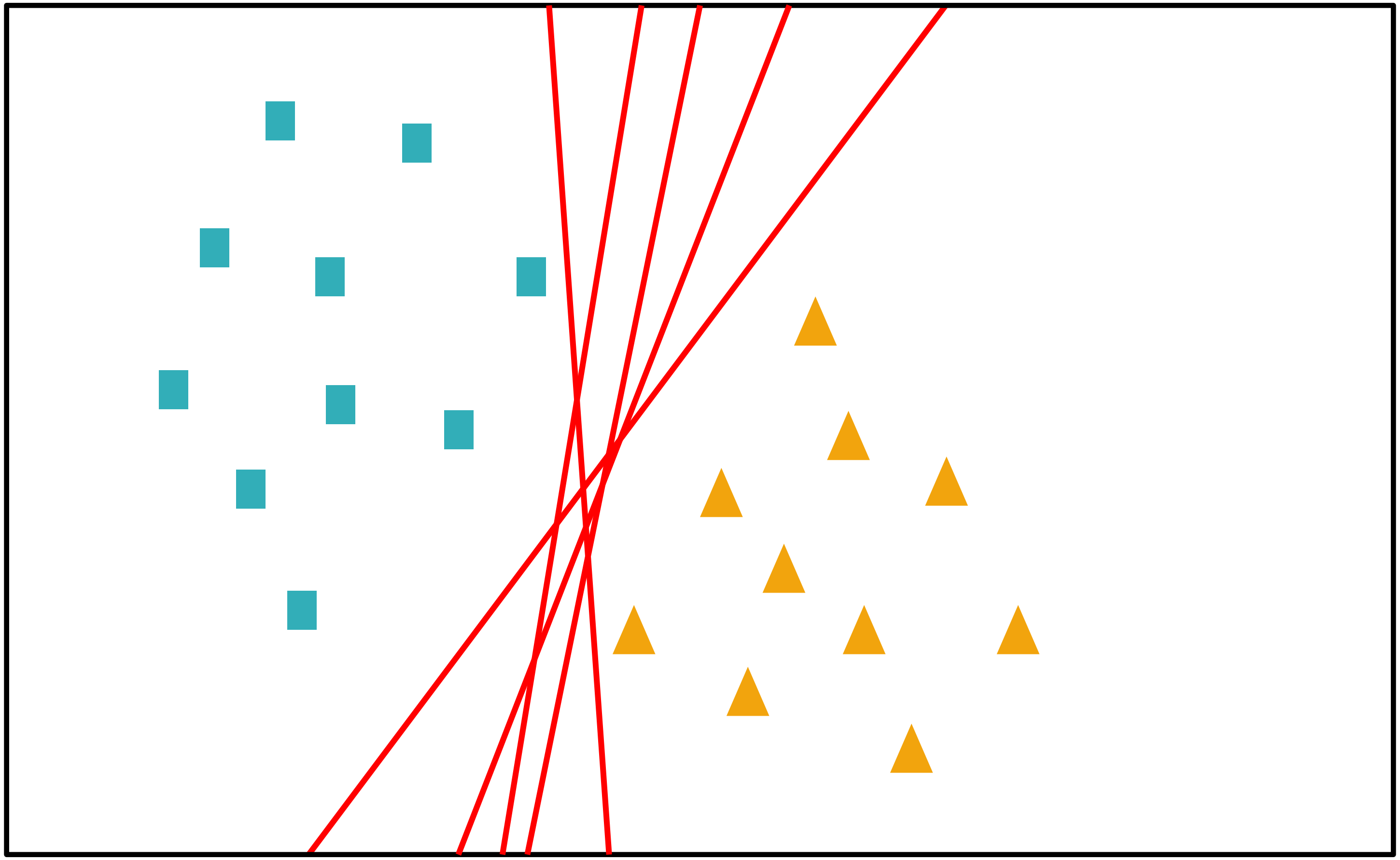
Why using SVM?

- High-Dimensionality
- Memory Efficiency
- Versatility

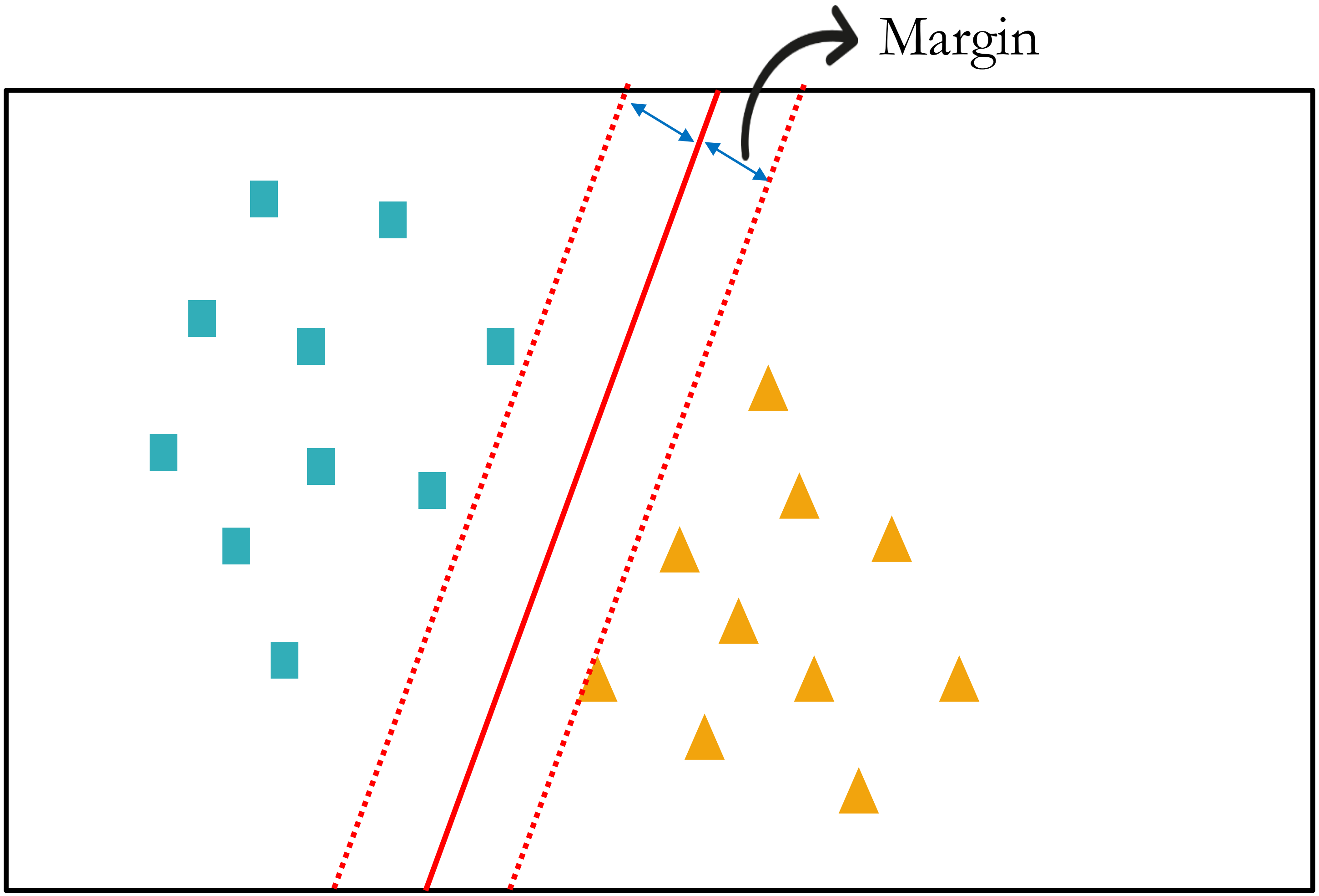
How SVM works



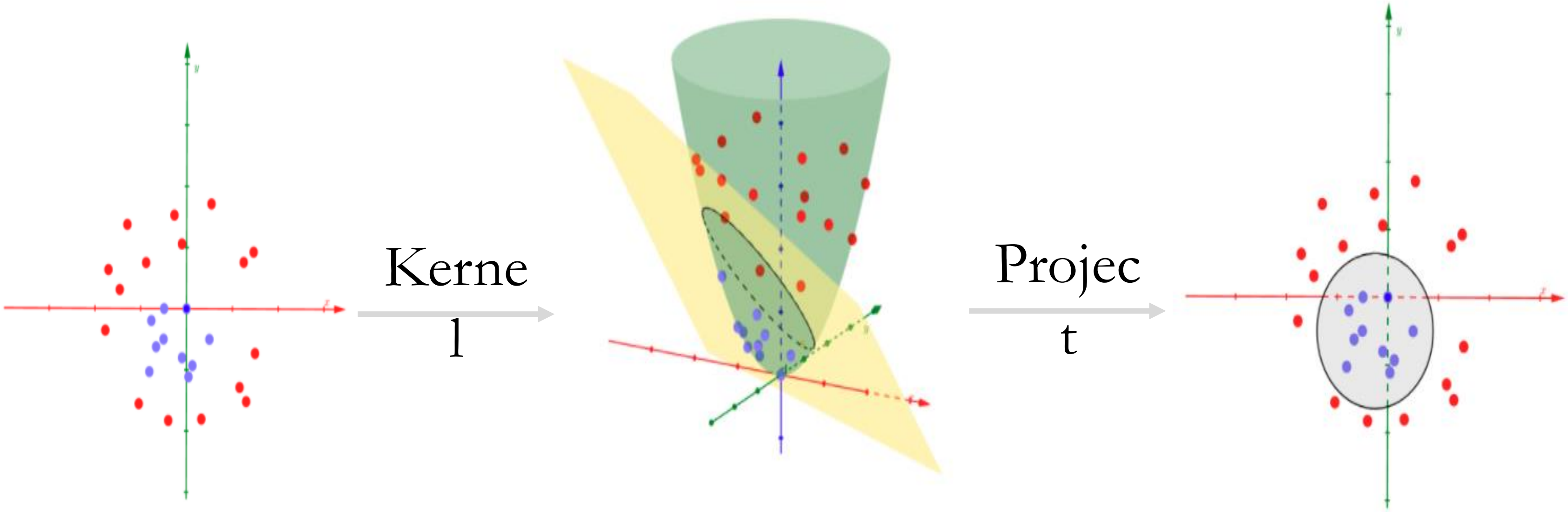
How SVM works



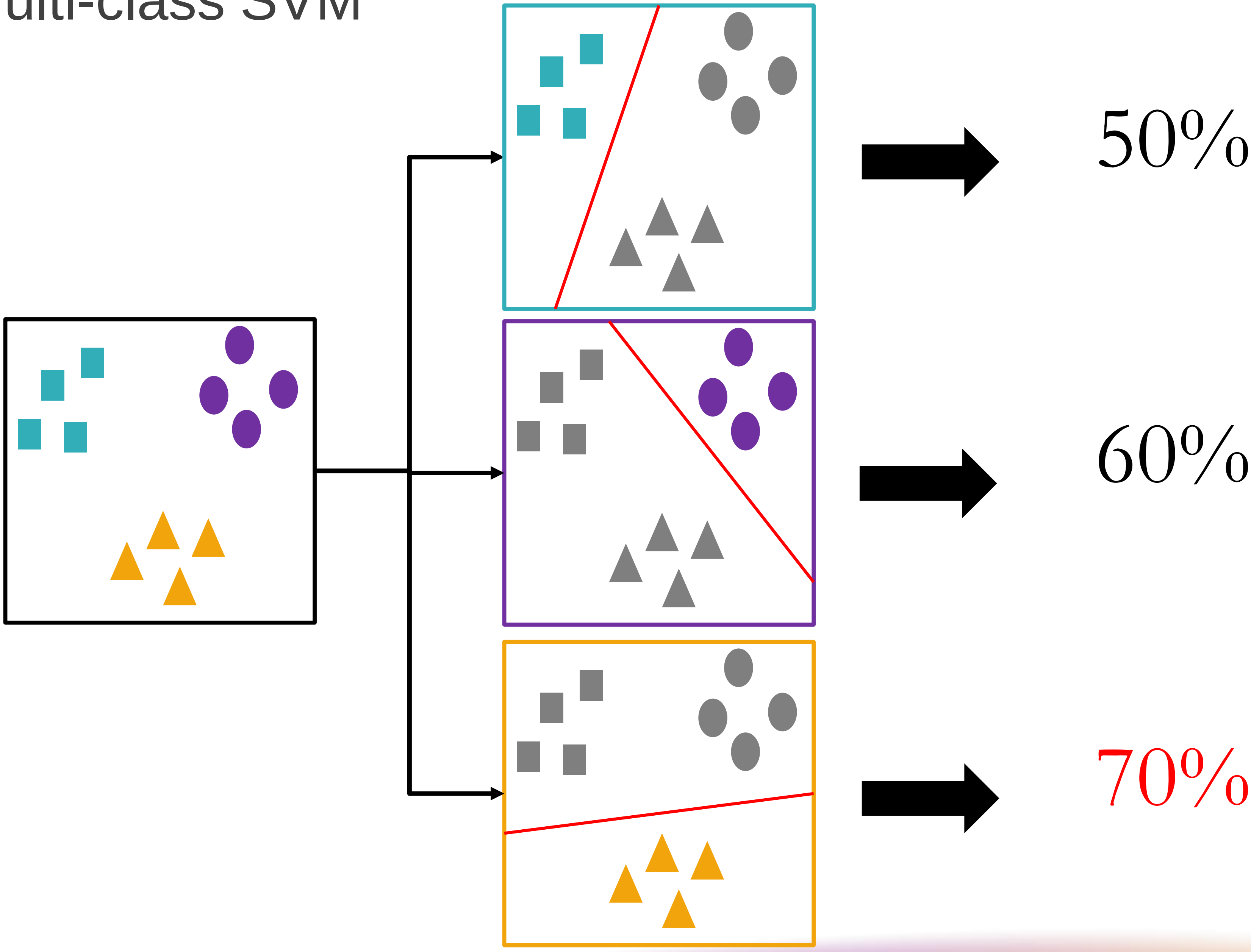
How SVM works



How SVM works



Multi-class SVM



Finding the hyperlane

Hyperplane: $w^T x + b = 0$

$$margin = \min_n \frac{y_n(w^T x_n + b)}{\|w\|_2}$$

Finding w and b:

$$(w, b) = \arg\max_{w,b} \left\{ \min_n \frac{y_n(w^T x_n + b)}{\|w\|_2} \right\} \rightarrow \boxed{(w, b) = \arg\min_{w,b} \frac{1}{2} \|w\|_2^2}$$

Quadratic Programming Solver

Problem: $\boldsymbol{\lambda} = \boldsymbol{arg}\ \mathbf{min}_{\lambda} \frac{\mathbf{1}}{\mathbf{2}} \boldsymbol{\lambda}^{\boldsymbol{T}} \boldsymbol{K} \boldsymbol{\lambda} + \boldsymbol{p}^{\boldsymbol{T}} \boldsymbol{\lambda}$

Subject to: $\boldsymbol{G\lambda} \leq \mathbf{h}$

$\boldsymbol{A\lambda} = \mathbf{b}$

- $\boldsymbol{\lambda}$ is Lagrange multiplier
  - $\boldsymbol{K}$ is square matrix
- $\boldsymbol{G}, \boldsymbol{A} \in R^{m \times n}$; $\boldsymbol{h}, \boldsymbol{b} \in R^{m}$; $\boldsymbol{p} \in R^{n}$

Solve problem using CVXOPT

$$\boldsymbol{w} = \sum_{n=1}^{N} \boldsymbol{\lambda}_n\, y_n x_n \longrightarrow \mathbf{b}$$

Build K

Create p

Build G

Build h

Build A

Build b

Find λ

Find w, b

$$\lambda = \arg\min_{\lambda} \frac{1}{2}\lambda^T K\lambda + p^T\lambda$$

Subject to: $G\lambda \leq \mathbf{h}$

$A\lambda = \mathbf{b}$

Training data

X =[[2.3, 1.7] y = [1
[1.5, 1.4] 1
[2.4, 2.1] 1
[3.4, 0.7] -1
[4.2 2.3]] -1]

$V = [y_1 X_1 + y_2 X_2 + \ldots + y_5 X_5]$ → V = [[2.3, 1.7]
[1.5, 1.4]
[2.4, 2.1]
[-3.4, -0.7]
[-4.2, -2.3]]

K = $V^T V$ → K=

$$\begin{bmatrix} 8.18 & 5.83 & 9.09 & -9.01 & -13.57 \\ 5.83 & 4.21 & 6.5 & -6.08 & -9.52 \\ 9.09 & 6.54 & 10.17 & -9.63 & -14.91 \\ -9.01 & -6.08 & -9.63 & 12.05 & 15.9 \\ -13.57 & -9.52 & -14.91 & 15.9 & 22.93 \end{bmatrix}$$

Build K

Create p

Build G

Build h

Build A

Build b

Find λ

Find w, b

$$\boldsymbol{\lambda} = \boldsymbol{arg}\ \min_{\boldsymbol{\lambda}} \frac{\mathbf{1}}{\mathbf{2}} \boldsymbol{\lambda}^{T} \boldsymbol{K} \boldsymbol{\lambda} + \boldsymbol{p}^{T} \boldsymbol{\lambda}$$

Subject to: $\boldsymbol{G\lambda} \leq \mathbf{h}$

$\boldsymbol{A\lambda} = \mathbf{b}$

Training data

X =[[2.3, 1.7]
[1.5, 1.4]
[2.4, 2.1]
[3.4, 0.7]
[4.2 2.3]]

y = [1
1
1
-1
-1]

p is all-one vector:

$$p = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

Build K

Create p

Build G

Build h

Build A

Build b

Find λ

Find w, b

$$\lambda = arg\min_{\lambda} \frac{1}{2}\lambda^T K \lambda + p^T \lambda$$

Subject to: $G\lambda \le \mathbf{h}$

$A\lambda = \mathbf{b}$

Training data

X =[[2.3, 1.7]
[1.5, 1.4]
[2.4, 2.1]
[3.4, 0.7]
[4.2 2.3]]

y = [1
1
1
-1
-1]

G is a diag matrix:

For all λ >= 0

$$G = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Build K

Create p

Build G

Build h

Build A

Build b

Find λ

Find w, b

$$\lambda = arg\ \min_{\lambda} \frac{1}{2} \lambda^T K \lambda + p^T \lambda$$

Subject to: $\boldsymbol{G\lambda} \leq \mathbf{h}$

$\boldsymbol{A\lambda} = \mathbf{b}$

Training data

X =[[2.3, 1.7]
[1.5, 1.4]
[2.4, 2.1]
[3.4, 0.7]
[4.2 2.3]]

y = [1
1
1
-1
-1]

$$h = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Build K

Create p

Build G

Build h

Build A

Build b

Find λ

Find w, b

$$\boldsymbol{\lambda} = \boldsymbol{arg}\ \min_{\lambda} \frac{1}{2} \boldsymbol{\lambda}^T \boldsymbol{K} \boldsymbol{\lambda} + \boldsymbol{p}^T \boldsymbol{\lambda}$$

Subject to: $\boldsymbol{G\lambda} \leq \mathbf{h}$

$\boldsymbol{A\lambda} = \mathbf{b}$

Training data

X =[[2.3, 1.7]
[1.5, 1.4]
[2.4, 2.1]
[3.4, 0.7]
[4.2 2.3]]

y = [1
1
1
-1
-1]

$$\text{A} = y^T = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 \end{bmatrix}$$

Build K

Create p

Build G

Build h

Build A

Build b

Find λ

Find w, b

$$\lambda = \arg\min_{\lambda} \frac{1}{2} \lambda^T K \lambda + p^T \lambda$$

Subject to: $G\lambda \leq \mathbf{h}$

$A\lambda = \mathbf{b}$

Training data

X =[[2.3, 1.7]
[1.5, 1.4]
[2.4, 2.1]
[3.4, 0.7]
[4.2 2.3]]

y = [1
1
1
-1
-1]

b =[0]

Build K

Create p

Build G

Build h

Build A

Build b

Find λ

Find w, b

$$\boldsymbol{\lambda} = \boldsymbol{arg}\,\min_{\lambda} \frac{1}{2}\boldsymbol{\lambda^T K \lambda} + \boldsymbol{p^T \lambda}$$

Subject to: $\boldsymbol{G\lambda} \leq \mathbf{h}$

$\boldsymbol{A\lambda} = \mathbf{b}$

Training data

X =[[2.3, 1.7] y = [1
[1.5, 1.4] 1
[2.4, 2.1] 1
[3.4, 0.7] -1
[4.2 2.3]] -1]

Apply K, p, G, h, A, b to below equation:

$$\boldsymbol{\lambda} = \begin{bmatrix} 9.61127142e-02 \\ 2.92683333e-09 \\ 8.97157531e-01 \\ 2.68704898e-08 \\ 9.93270222e-01 \end{bmatrix}$$

Build K

Create p

Build G

Build h

Build A

Build b

Find λ

Find w, b

$$\lambda = \arg\min_{\lambda} \frac{1}{2}\lambda^T K \lambda + p^T \lambda$$

Subject to: $\boldsymbol{G\lambda} \leq \mathbf{h}$

$\boldsymbol{A\lambda} = \mathbf{b}$

Training data

X =[[2.3, 1.7]   y = [1
[1.5, 1.4]   1
[2.4, 2.1]   1
[3.4, 0.7]   -1
[4.2 2.3]]   -1]

Apply $\boldsymbol{\lambda}$, X, y to $\boldsymbol{w} = \sum_{n=1}^{N} \lambda_n y_n x_n$

w = [-1.34287034, 0.42806505]

b = 3.45

Support Vector Machine (Classification)

- Pima dataset (ratio 6:4)
- SVM (sklearn)
- Evaluation: Accuracy

Time: 15 minutes

SVM (Regresion)

Time: 15 minutes