Anomaly Detection Using Support Vector Machines

COMP90073
Security Analytics

Sarah Erfani, CIS

Semester 2, 2021

Outline

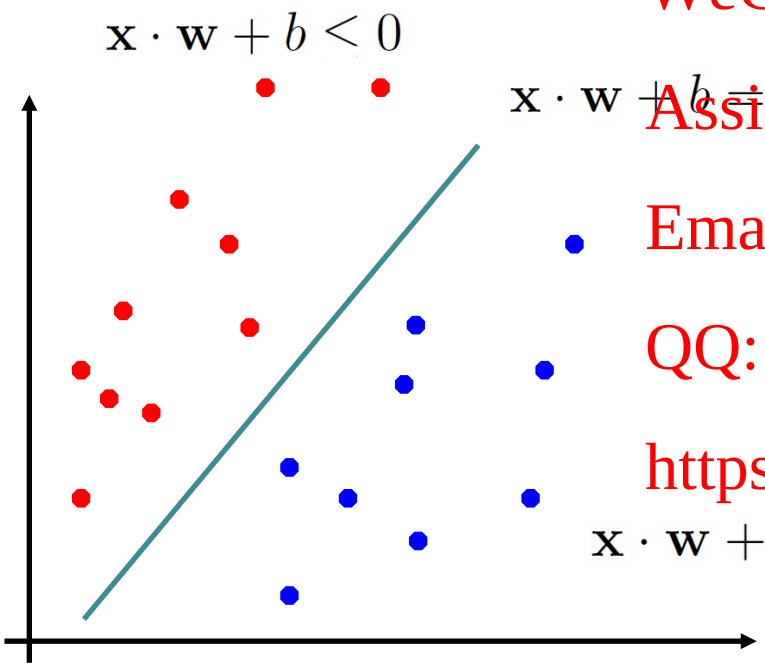
- Review of SVM
- Support Vector Data Desc /DD)
- One-class Support Vector Machine (OCSVM)
- Recent developments of OCSVM/SVDD

SVM – Revision

$\{\mathbf{x}_i, y_i\}$ where $i = 1 \ldots L, y_i \in \{-1, 1\}, \mathbf{x}_i \in \mathbb{R}^D$

This hyperplane can be descri $\mathbf{w} + b = 0$ where:

- $\mathbf{w}$ is normal to the hype
- $\frac{b}{\|\mathbf{w}\|}$ is the perpendicular distance from the hyperplane to the origin.



Classification rule

$$f(\mathbf{x}) = \text{sign}(\mathbf{x} \cdot \mathbf{w} + b) = \begin{cases} +1 \text{ if } \mathbf{x} \cdot \mathbf{w} + b \geq 0 \\ -1 \text{ if } \mathbf{x} \cdot \mathbf{w} + b < 0 \end{cases}$$

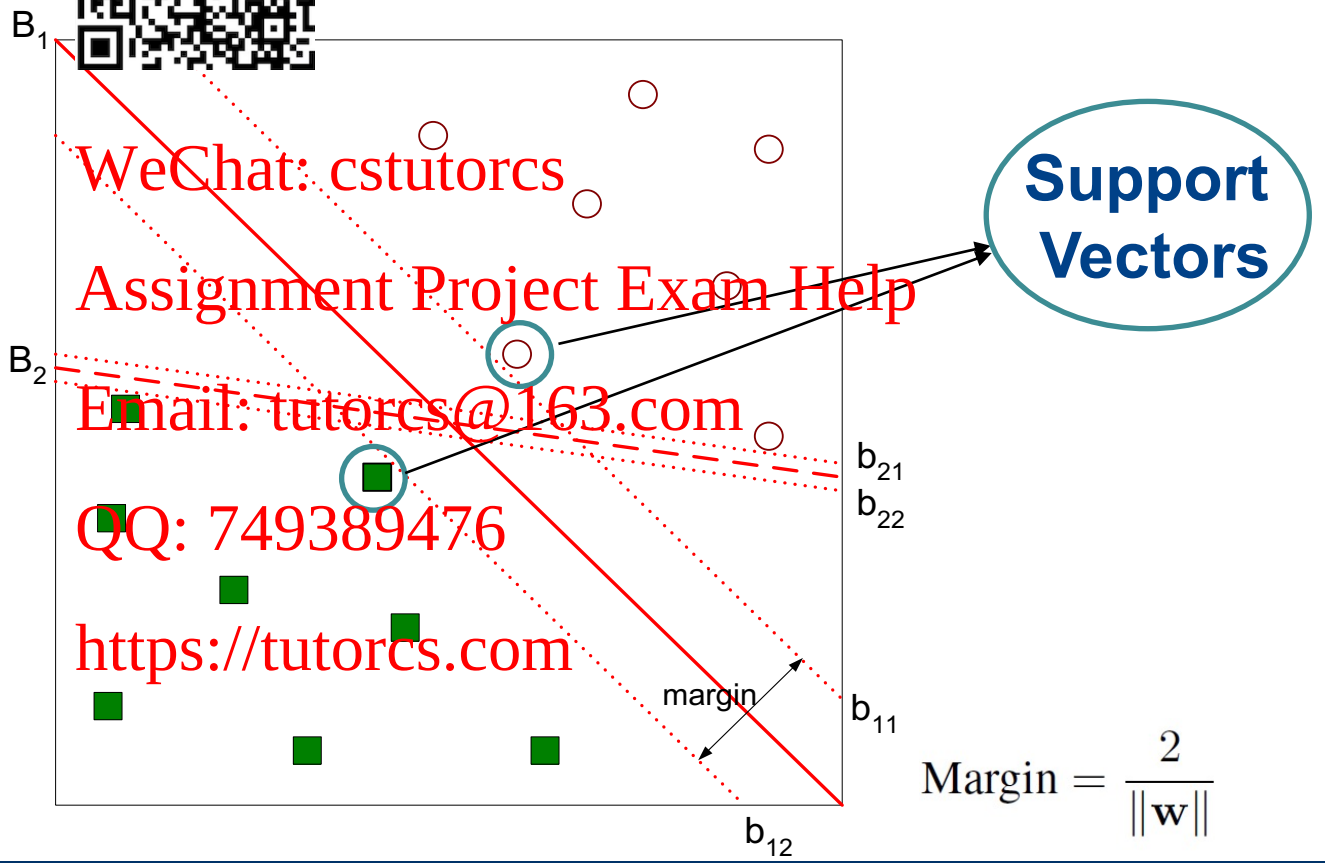
Find $\mathbf{w}$ and b such that:

$$\mathbf{x}_i \cdot \mathbf{w} + b \geq 0 \text{ for } y_i = +1$$
$$\mathbf{x}_i \cdot \mathbf{w} + b < 0 \text{ for } y_i = -1$$
$$\text{for all } i = 1 \ldots L$$

Training objective

Large Margin Classifiers – Revision

- Find hyperplane maximises the margin => B1 is better than B2
- Margin: sum of shortest di[QR]om the planes to the positive/negative samples



Solving the Optimization Problem – Revision

***Primal problem**: solve for **w** and b*

$$\min \frac{1}{2}\|w\|^2 \qquad \text{s.t.} \; y_i(w^T\phi(x_i)+b) \geq 1, \;\; \forall\, i = 1, \ldots, n$$

*Equivalent **dual problem** formulation: solve for $\alpha_1 \ldots \alpha_L$*: Lagrange multipliers for each data point

$$\max_{\alpha} \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{L} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

s.t.

$$\alpha_i \geq 0$$

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

More convenient to solve

Soft Margin Classification – Revision

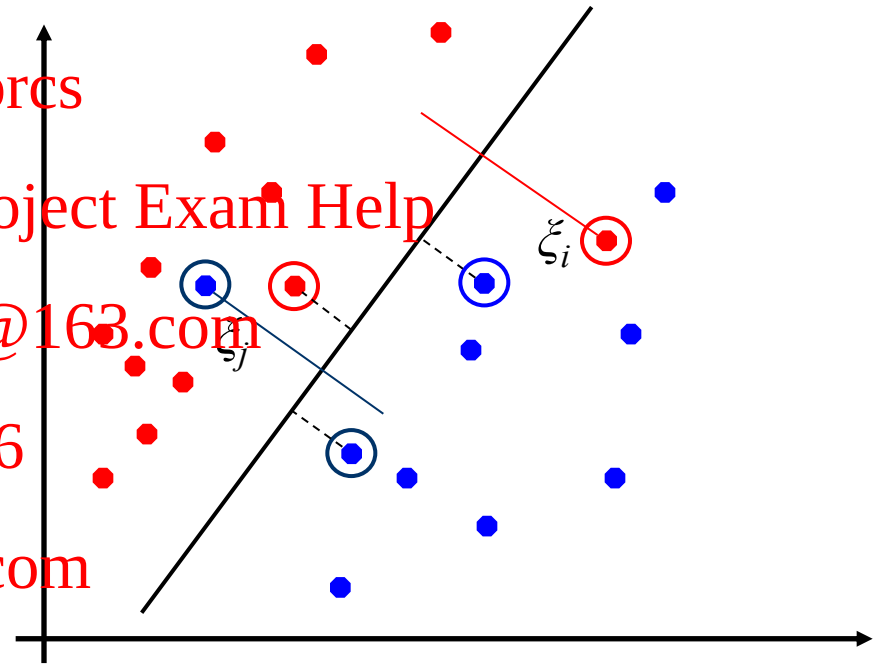
- ***Slack variables*** ξ_i can be added to allow misclassification of difficult or noisy examples, resulting marg[illegible]*oft*.

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i$$

s.t.

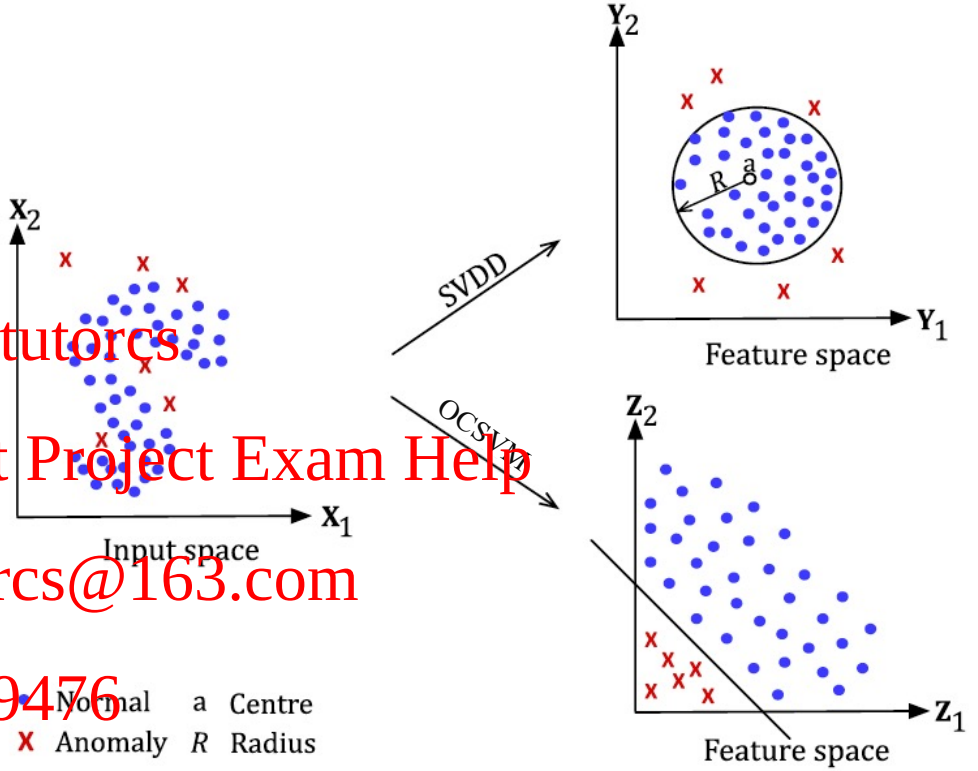
$$y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i, \forall i = 1, \ldots, n$$

$$\xi_i \geq 0, \forall\, i = 1, \ldots, n$$



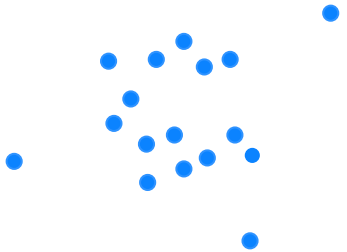
Anomaly Detection using SVM

- **Assumption:** All (or majority of) training examples belong to the *normal* (positive) class.
- **Objective:** identify anomalies by modeling normal pattern
  - Support Vector Data Description (SVDD) [1]
  - One-class Support Vector Machine (OCSVM) [2]



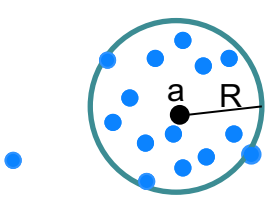
Support Vector Data Description (SVDD) [1]

- Find the minimal circumscribing hyperball in high-dimensional space encom[illegible]ost) all the observations.



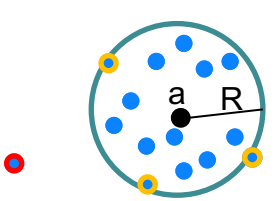
Support Vector Data Description (SVDD) [1]

- Find the minimal circumscribing hyperball in high-dimensional space encom[illegible]ost) all the observations.



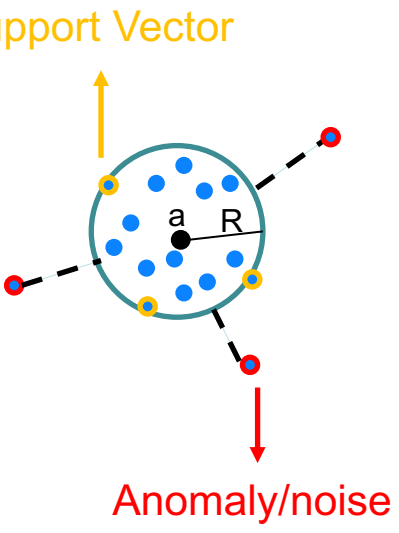
Support Vector Data Description (SVDD)

- Find the minimal circumscribing hyperball in high-dimensional space encom[illegible]ost) all the observations.



Support Vector Data Description (SVDD)

- Find the minimal circumscribing hyperball in high-dimensional space encom[illegible]ost) all the observations



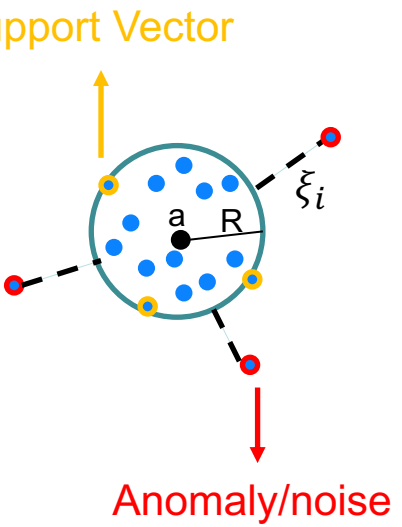
Support Vector Data Description (SVDD)

- Find the minimal circumscribing hyperball in high-dimensional space encom[illegible]ost) all the observations

$$\min_{R,\xi,a} R^2 + \sum_{i=1}^{n} \xi_i$$

s.t.

$$\|\phi(x_i) - a\|^2 \leq R^2 + \xi_i, \forall\, i = 1, \ldots, n$$
$$\xi_i \geq 0, \forall\, i = 1, \ldots, n$$



where,

- R: Reduce of the ball
- ξ: Slack variable
- a: Center of the ball
- $\phi(.)$: non-linear function

Support Vector Data Description (SVDD)

Lagrangian form:

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\right) - \sum_{i=1}^{n}\gamma_i\xi_i$$

where, $\gamma_i \geq 0$ and $\alpha_i \geq 0$ are Lagrange multipliers.

Support Vector Data Description (SVDD)

Lagrangian form:

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\right) - \sum_{i=1}^{n}\gamma_i\xi_i$$

where, $\gamma_i \geq 0$ and $\alpha_i \geq 0$ are Lagrange multipliers.

Set the derivatives with respect to the primal variables $R, a, \boldsymbol{\xi}$ equal to zero, we get

- $\frac{\partial L}{\partial R} = ?$
- $\frac{\partial L}{\partial a} = ?$
- $\frac{\partial L}{\partial \xi_i} = ?$

Support Vector Data Description (SVDD)

Lagrangian form:

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\right) - \sum_{i=1}^{n}\gamma_i\xi_i$$

where, $\gamma_i \geq 0$ and $\alpha_i \geq 0$ are Lagrange multipliers.

Set the derivatives with respect to the primal variables $R, a, \boldsymbol{\xi}$ equal to zero, we get

- $\frac{\partial L}{\partial R} = 2R - 2R\sum_{i=1}^{n}\alpha_i = 0$ $\quad\Rightarrow\quad \sum_{i=1}^{n}\alpha_i = 1$
- $\frac{\partial L}{\partial a} = ?$
- $\frac{\partial L}{\partial \xi_i} = ?$

Support Vector Data Description (SVDD)

Lagrangian form:

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\right) - \sum_{i=1}^{n}\gamma_i\xi_i$$

where, $\gamma_i \geq 0$ and $\alpha_i \geq 0$ are Lagrange multipliers.

Set the derivatives with respect to the primal variables $R, a, \boldsymbol{\xi}$ equal to zero, we get

- $\frac{\partial L}{\partial R} = 2R - 2R\sum_{i=1}^{n}\alpha_i = 0$ $\quad\Rightarrow\quad \sum_{i=1}^{n}\alpha_i = 1$
- $\frac{\partial L}{\partial a} = 2a\sum_{i=1}^{n}\alpha_i - 2\sum_{i=1}^{n}\alpha_i\phi(x_i) = 0$ $\quad\Rightarrow\quad a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
- $\frac{\partial L}{\partial \xi_i} = ?$

Support Vector Data Description (SVDD)

Lagrangian form:

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\right) - \sum_{i=1}^{n}\gamma_i\xi_i$$

where, $\gamma_i \geq 0$ and $\alpha_i \geq 0$ are Lagrange multipliers.

Set the derivatives with respect to the primal variables $R, a, \boldsymbol{\xi}$ equal to zero, we get

- $\frac{\partial L}{\partial R} = 2R - 2R\sum_{i=1}^{n}\alpha_i = 0$ $\quad\Rightarrow\quad \sum_{i=1}^{n}\alpha_i = 1$

- $\frac{\partial L}{\partial a} = 2a\sum_{i=1}^{n}\alpha_i - 2\sum_{i=1}^{n}\alpha_i\phi(x_i) = 0$ $\quad\Rightarrow\quad a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$

- $\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \gamma_i = 0$ $\quad\Rightarrow\quad C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$
$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\right) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$\sum_{i=1}^{n}\alpha_i = 1$
$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
$C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\Big(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\Big) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i$$

$$+ \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - 2\sum_{i=1}^{n}\alpha_i\,\phi(x_i)^T a + a^T a\sum_{i=1}^{n}\alpha_i - R^2\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}(\alpha_i+\gamma_i)\xi_i$$

$\sum_{i=1}^{n}\alpha_i = 1$
$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
$C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$L(a, R, \xi, \alpha, \gamma)$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left((\phi(x_i) - a)^T(\phi(x_i) - a)\right) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - 2\sum_{i=1}^{n}\alpha_i\,\phi(x_i)^T a + a^T a\sum_{i=1}^{n}\alpha_i - R^2\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}(\alpha_i+\gamma_i)\xi_i$$

$\sum_{i=1}^{n}\alpha_i = 1$
$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
$C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\Big(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\Big) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i$$

$$+ \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - 2\sum_{i=1}^{n}\alpha_i\,\phi(x_i)^T a + a^T a\sum_{i=1}^{n}\alpha_i - R^2\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}(\alpha_i+\gamma_i)\xi_i$$

$\sum_{i=1}^{n}\alpha_i = 1$
$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
$C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\Big(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\Big) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i$$

$$+ \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - 2\sum_{i=1}^{n}\alpha_i\,\phi(x_i)^T a + a^T a\sum_{i=1}^{n}\alpha_i - R^2\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}(\alpha_i+\gamma_i)\xi_i$$

$\sum_{i=1}^{n}\alpha_i = 1$
$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
$C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\Big(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\Big) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i$$

$$+ \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - 2\sum_{i=1}^{n}\alpha_i\,\phi(x_i)^T a + a^T a\sum_{i=1}^{n}\alpha_i - R^2\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}(\alpha_i+\gamma_i)\xi_i$$

$$\sum_{i=1}^{n}\alpha_i = 1$$
$$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$$
$$C = \alpha_i + \gamma_i$$

Support Vector Data Description (SVDD)

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\Big(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\Big) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i$$

$$+ \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - 2\sum_{i=1}^{n}\alpha_i\,\phi(x_i)^T a + a^T a\sum_{i=1}^{n}\alpha_i - R^2\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}(\alpha_i+\gamma_i)\xi_i$$

$\sum_{i=1}^{n}\alpha_i = 1$
$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
$C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\Big(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\Big) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i$$

$$+ \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - 2\sum_{i=1}^{n}\alpha_i\,\phi(x_i)^T a + a^T a\sum_{i=1}^{n}\alpha_i - R^2\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}(\alpha_i+\gamma_i)\xi_i$$

=1 =1

$\sum_{i=1}^{n}\alpha_i = 1$
$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
$C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\Big(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\Big) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i$$

$$+ \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - 2\sum_{i=1}^{n}\alpha_i\,\phi(x_i)^T a + a^T a\sum_{i=1}^{n}\alpha_i - R^2\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}(\alpha_i+\gamma_i)\xi_i$$

$\sum_{i=1}^{n}\alpha_i$ =1 ; $\sum_{i=1}^{n}\alpha_i$ =1 ; $(\alpha_i+\gamma_i)$ =C

$\sum_{i=1}^{n}\alpha_i = 1$
$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
$C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\Big(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\Big) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i$$

$$+ \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - 2\sum_{i=1}^{n}\alpha_i\,\phi(x_i)^T a + a^T a\sum_{i=1}^{n}\alpha_i - R^2\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}(\alpha_i+\gamma_i)\xi_i$$

$=a^2$ $=1$ $=1$ $=C$

$\sum_{i=1}^{n}\alpha_i = 1$
$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
$C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\Big(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\Big) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i$$

$$+ \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - 2\sum_{i=1}^{n}\alpha_i\,\phi(x_i)^T a + a^T a\sum_{i=1}^{n}\alpha_i - R^2\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}(\alpha_i+\gamma_i)\xi_i$$

$=a^T$ $=1$ $=1$ $=C$

$$= \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - a^T a$$

$\sum_{i=1}^{n}\alpha_i = 1$
$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
$C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$L(a, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\Big(R^2 + \xi_i - (\phi(x_i) - a)^T(\phi(x_i) - a)\Big) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{n}\xi_i$$

$$+ \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - 2\sum_{i=1}^{n}\alpha_i\,\phi(x_i)^T a + a^T a\sum_{i=1}^{n}\alpha_i - R^2\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}(\alpha_i+\gamma_i)\xi_i$$

$=a^T$ $=1$ $=1$ $=C$

$$= \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - a^T a$$

$$= \sum_{i=1}^{n}\alpha_i k(x_i, x_i) - \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j k(x_i, x_j)$$

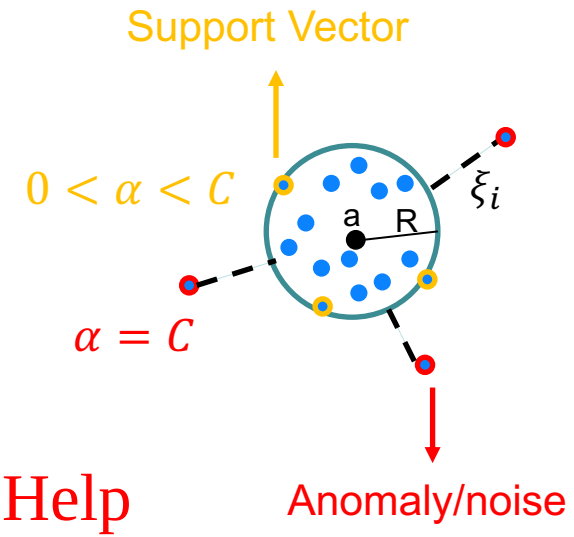
$\sum_{i=1}^{n}\alpha_i = 1$
$a = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
$C = \alpha_i + \gamma_i$

Support Vector Data Description (SVDD)

$$\underset{\alpha}{\text{argmax}} \sum_{i=1}^{n} \alpha_i k(x_i, x_i) - \alpha_j k(x_i, x_j)$$

s.t.

$$0 \leq \alpha_i \leq C, \qquad \sum_{i=1}^{n} \alpha_i = 1$$

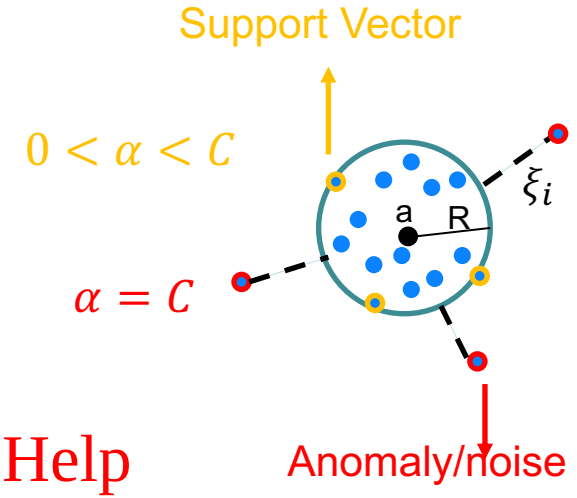


Support Vector Data Description (SVDD)

$$\underset{\alpha}{\text{argmax}} \sum_{i=1}^{n} \alpha_i k(x_i, x_i) - \alpha_j k(x_i, x_j)$$

s.t.

$$0 \le \alpha_i \le C, \quad \sum_{i=1}^{n} \alpha_i = 1$$



- $\|\phi(x_i) - a\|^2 < R^2 \rightarrow \alpha_i = 0$
- $\|\phi(x_i) - a\|^2 = R^2 \rightarrow 0 < \alpha_i < C$
- $\|\phi(x_i) - a\|^2 > R^2 \rightarrow \alpha_i = C$

Support Vector Data Description (SVDD)

- $a = \sum_{i=1}^{n} \alpha_i \phi(x_i)$

Support Vector Data Description (SVDD)

- $a = \sum_{i=1}^{n} \alpha_i \phi(x_i)$
- $R^2 = \|\phi(x_i) - a\|^2$, where x_i is any support vectors with $0 < \alpha_i < C$

$$R^2 = k(x_i, x_i) - 2\sum_{i=1}^{n} \alpha_j k(x_i, x_j) + \sum_{i=1}^{n}\sum_{i=1}^{n} \alpha_i \alpha_j k(x_i, x_j)$$

Support Vector Data Description (SVDD)

- $a = \sum_{i=1}^{n} \alpha_i \phi(x_i)$
- $R^2 = \|\phi(x_i) - a\|^2$, where x_i is any support vectors with $0 < \alpha_i < C$

$$R^2 = k(x_i, x_i) - 2\sum_{i=1}^{n} \alpha_j k(x_i, x_j) + \sum_{i=1}^{n}\sum_{i=1}^{n} \alpha_i \alpha_j k(x_i, x_j)$$

- New sample z is identified as normal if

$$\|z - a\|^2 = (z.z) - 2\sum_{i=1}^{n} \alpha_j (z.x_i) + \sum_{i=1}^{n}\sum_{i=1}^{n} \alpha_i \alpha_j (x_i.x_j) \leq R^2$$

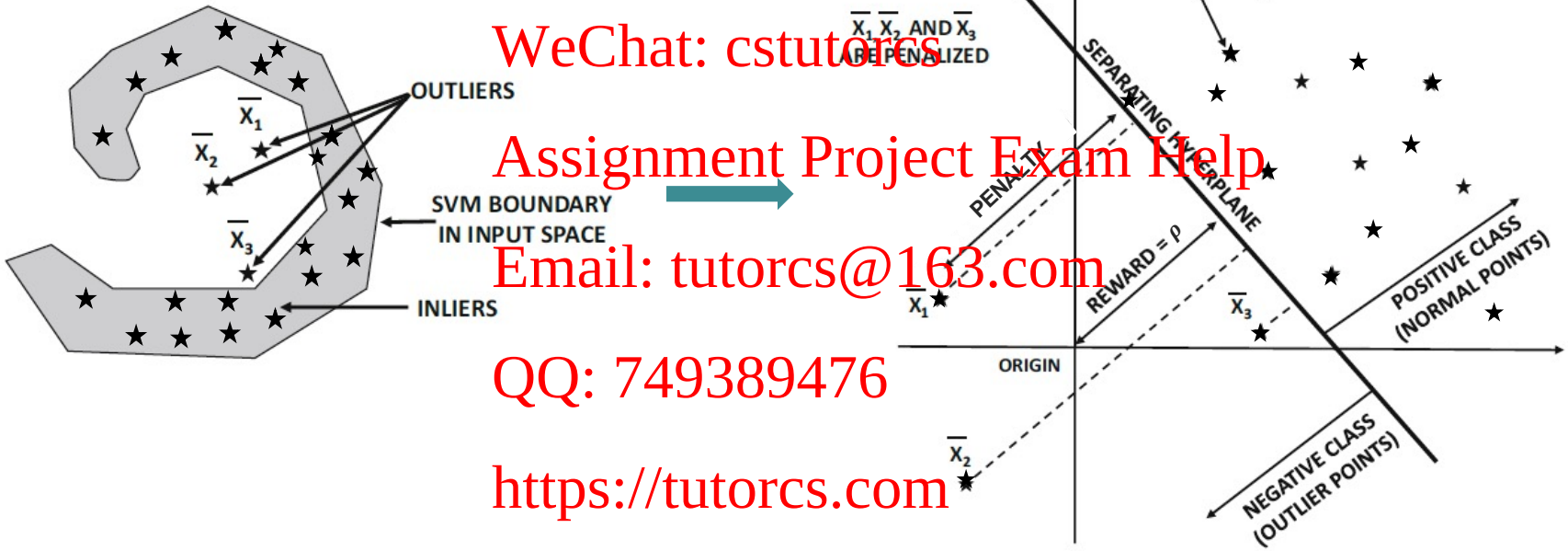
One-class Support Vector Machine (OCSVM) [2]

- Map input data into a high dimensional feature space
- Iteratively finds the maxim  in the hyperplane which best separates the training *data from the orig*
- Decision boundary

$$w \cdot \phi(x_i) - \rho = 0$$

- Formulate the optimization problem so it returns positive for as many of the N training examples as possible.

OCSVM Intuition



OCSVM

- Solve quadratic problem

$$\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho$$

s.t.

$$(w \cdot \phi(x_i)) \geq \rho - \xi_i, \forall\, i = 1, \ldots, n$$

$$\xi_i \geq 0, \forall\, i = 1, \ldots, n$$

- ν:
  - $\nu = \frac{1}{C}$
  - A prior probability that a data point in the training set is an anomaly.
  - Regulates the trade-off between false positives and false negatives in this model.
  - Due to the importance of ν, OCSVM is often referred to as ν-SVM

One-class Support Vector Machine (OCSVM)

- The problem can be simplified to

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j k(x_i, x_j)$$

$$\text{s.t.} \quad 0 \le \alpha_i \le \frac{1}{\nu n} \qquad \sum_{i=1}^{n}\alpha_i = 1$$

One-class Support Vector Machine (OCSVM)

- The problem can be simplified to

$$\min_{\boldsymbol{\alpha}} \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j)$$

$$\text{s.t.} \quad 0 \le \alpha_i \le \frac{1}{\nu n}, \quad \sum_{i=1}^{n} \alpha_i = 1$$

- Soft margin SVM for binary classification:

$$\operatorname{argmin}_{\alpha} \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j k(x_i, x_j) - \sum_{i=1}^{n} \alpha_i$$

$$\text{s.t.} \quad \alpha_i \ge 0, \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

One-class Support Vector Machine (OCSVM)

- The problem can be simplified to

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j k(x_i, x_j)$$

$$\text{s.t.} \quad 0 \le \alpha_i \le \frac{1}{\nu n}, \qquad \sum_{i=1}^{n}\alpha_i = 1$$

- $w = \sum_{i=1}^{n}\alpha_i\phi(x_i)$
- $\rho = w.\phi(x_i)$

One-class Support Vector Machine (OCSVM)

- Anomaly score for new sample z:
  - $Score(z) = \sum_{i=1}^{n} \alpha_i k($
  - $Score(z) < 0$: Anoma
  - $Score(z) \geq 0$: Normal

- For especial case of (invariant) kernels, e.g., Gaussian, $\text{SVDD} \cong \text{OCSVM}$

Deep SVDD [4]

Figure: Deep SVDD learns a neural network transformation $\phi(\cdot;\mathcal{W})$ with weights $\mathcal{W}$ from input space $\mathcal{X} \subseteq \mathbb{R}^d$ to output space $\mathcal{F} \subseteq \mathbb{R}^p$ that attempts to map most of the data network representations into a hypersphere characterized by centre c and radius R of minimum volume. Mappings of normal examples fall within, whereas mappings of anomalies fall outside the hypersphere.

Hybrid Deep-1SVM [5]

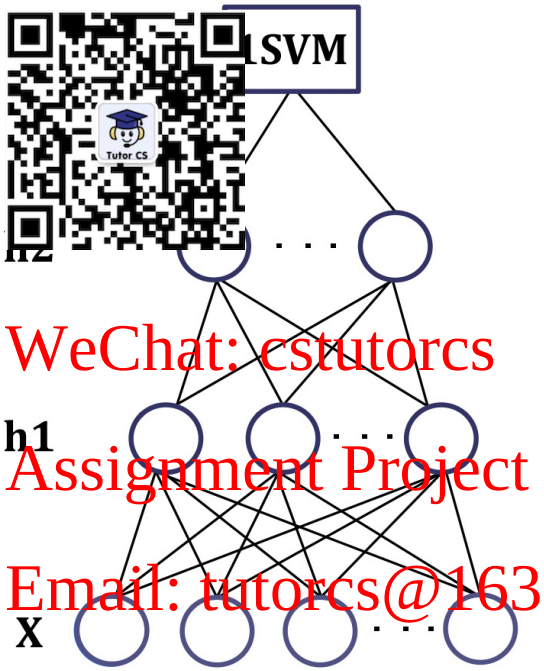


Figure: Deep model (AE) is trained to extract features that are relatively invariant to irrelevant variations in the input, so that the one-class SVM (1SVM) can effectively separate the normal data from anomalies in the learned feature space, using linear kernel.

Deep Learning and One-class SVM based Anomalous Crowd Detection [6]

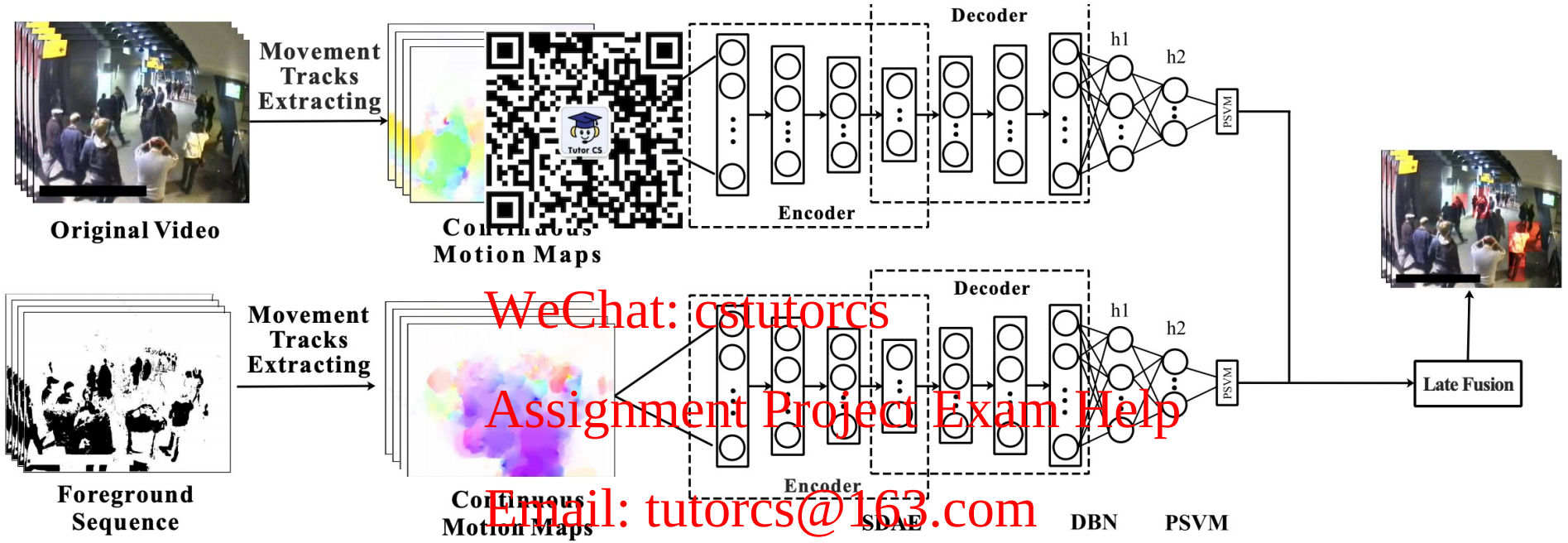


Figure: Original video and foreground video sequence are taken as the input of two branches of channel, then movement tracks are extracted to produce continuous motion maps. Training and testing on a hybrid deep learning model SDAE-DBN-PSVM, then follows to achieve anomalous event detection

Summary

- What is SVDD?
- How to derive dual formul /DD?
- How to extend SVM to OCSVM?

Next: Autoencoders and their applications

References

1. David M.J. Tax, Robert P.W. Duin, "Support Vector Data Description", Machine Learning, 2004

2. Bernhard Schölkopf, John C. Platt, John Shawe-Taylor, Alex J. Smola and Robert C. Williamson "Estimating the Support of a High-Dimensional Distribution". Neural Computation, 2001.

3. Charu C. Aggarwal, "Outlier Analysis", Springer, 2016. Chapter 3

4. Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. "Deep one-class classification." In *International Conference on Machine Learning (ICML)*, pp. 4393-4402. 2018.

5. Sarah Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie. "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning." *Pattern Recognition* 58 (2016): 121-134.

References

6. Meng Yang, Sutharshan Rajasegarar, Sarah M. Erfani, and Christopher Leckie. "Deep Learning and One-Class SVM based Anomalous Crowd Detection." In *IEEE International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8., 2019.