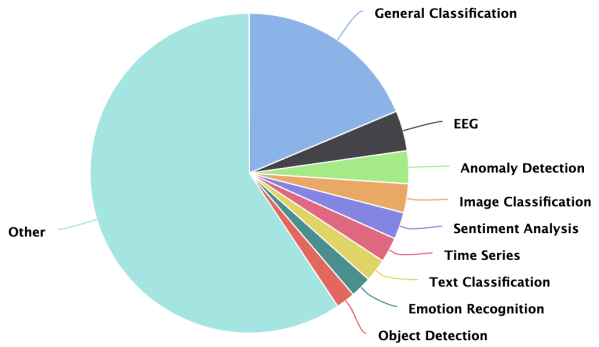
목차

SVM이 적용될 수 있는 다양한 분야들

Tasks

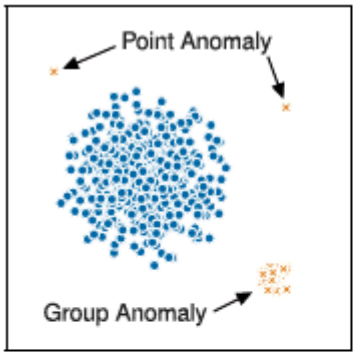


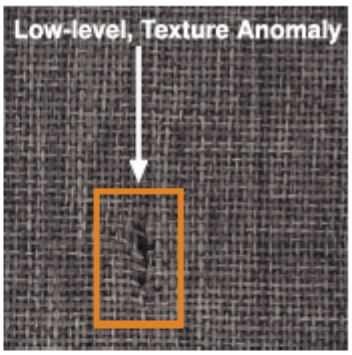
Task	Papers	Share
General Classification	96	18.68%
EEG	21	4.09%
Anomaly Detection	17	3.31%
Image Classification	15	2.92%
Sentiment Analysis	14	2.72%
Time Series	13	2.53%
Text Classification	12	2.33%
Emotion Recognition	11	2.14%
Object Detection	10	1.95%

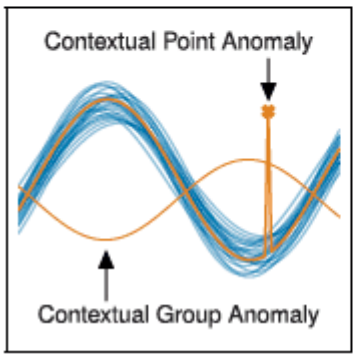
- General Classification : general task (image, text, ...)
- EEG(Electroencephalogram) : 뇌에 부착된 작은 금속 디스크(전극)를 사용하여 뇌의 전기적 활동을 측정하는 검사
- ...

Anomaly Detection

- 정상 범주에서 벗어나 있는 모든 상태, 물체 등을 감지하는 기술



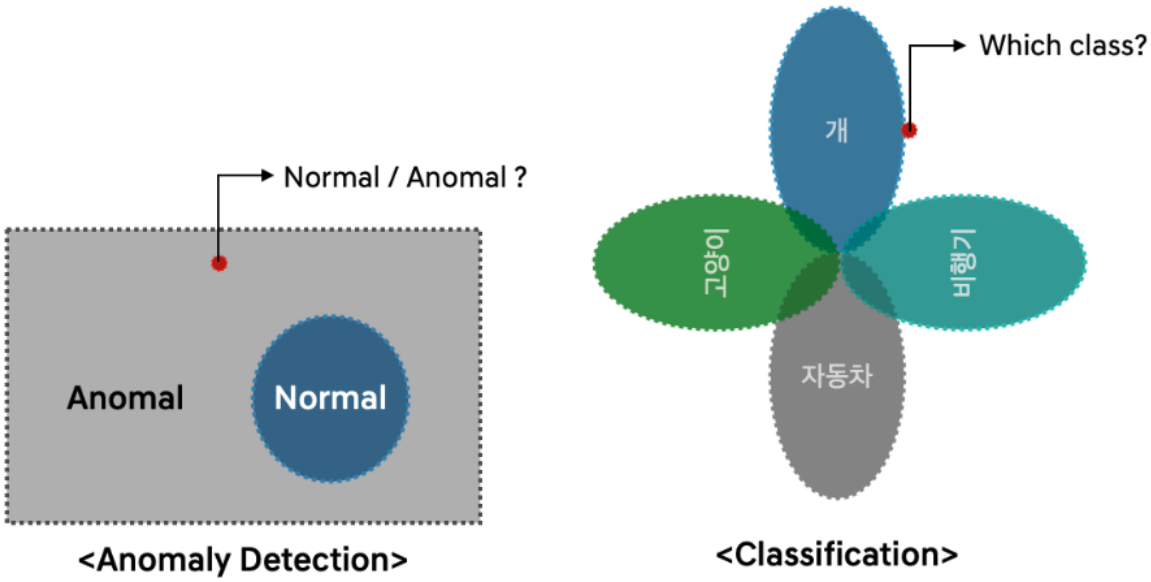




Anomaly Detection 의 종류

- Point anomaly : 정상 범주에서 벗어난 특정 사건, 거래, 이미지 등을 의미
    - 결제 시스템에서의 비정상적 거래, 제조 현장에서의 품질 불량 이미지 등
- Group anomaly : 하나의 Point만으로는 판단 불가하며, 다량의 정보로 판단 가능
    - 사이버 보안 등
- Contextual anomaly : Point anomaly와는 달리 한 점만으로는 판별 불가하며, Time series 내 문맥을 통해 파악해야함
    - 주식 시장에서의 비 이상적 과열 현상, 이상 기온 현상 등
- Low-level anomaly : Low-level feature는 semantic 하지 않은 noise 정보를 의미
    - 제조 환경에서의 품질검사 사례 등
- High-level anomaly : semantic 한 의미를 갖는 정보들을 다룸
    - 고양이 그림 사이에서 강아지가 등장했을 때 등
- Test Instance에 대해 Normal 또는 Anomaly로 구분해야된다는 관점에서 One Class Classification이라고도 할 수 있음

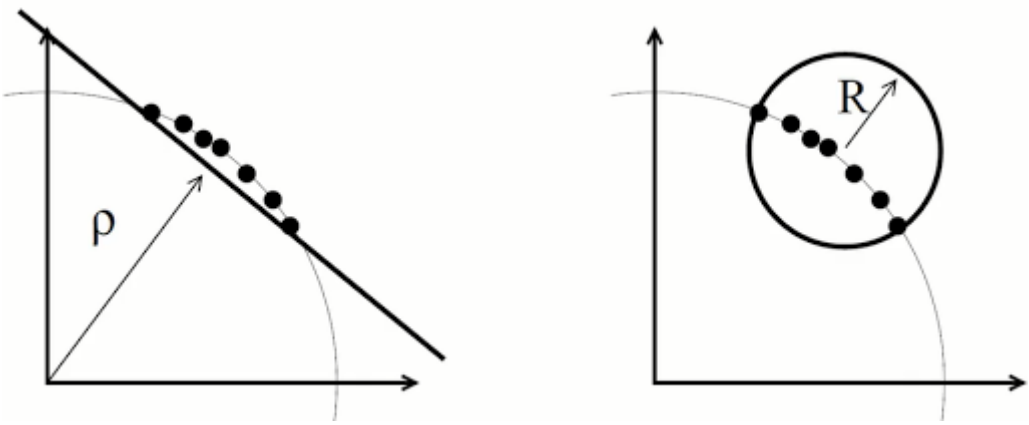
- Classification과 Anomaly Detection의 차이점

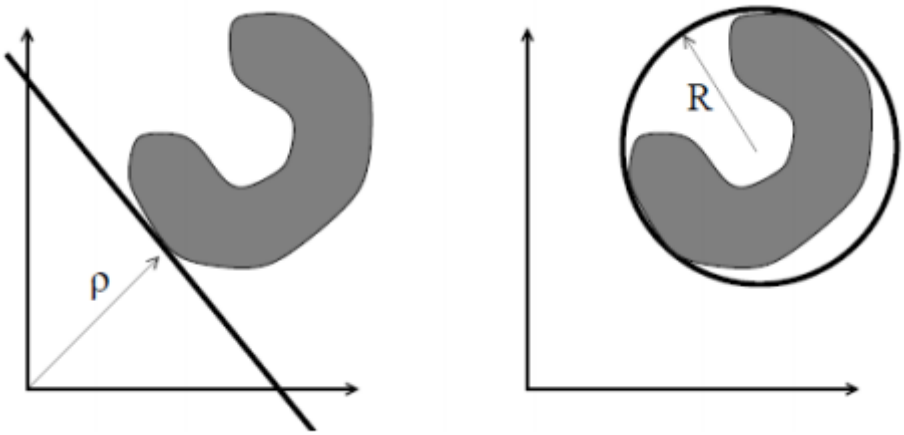


- **Classification** : 다양한 클래스의 데이터를 학습 단계에서 사용하며 Test Instance에 대해 학습 단계에서 본 클래스별 분포에서 어디에 가장 가까운지를 찾는 문제
- **Anomaly Detection** : 학습 단계에서는 Anomal Data를 볼 수 없음. 따라서 Normal Data 만을 학습 단계에서 사용하여 Normal Data 분포를 추정하여 Test Instance에 대해 Normal Data 분포와 얼나 다른지 계산함

SVDD (Support Vector Data Description)

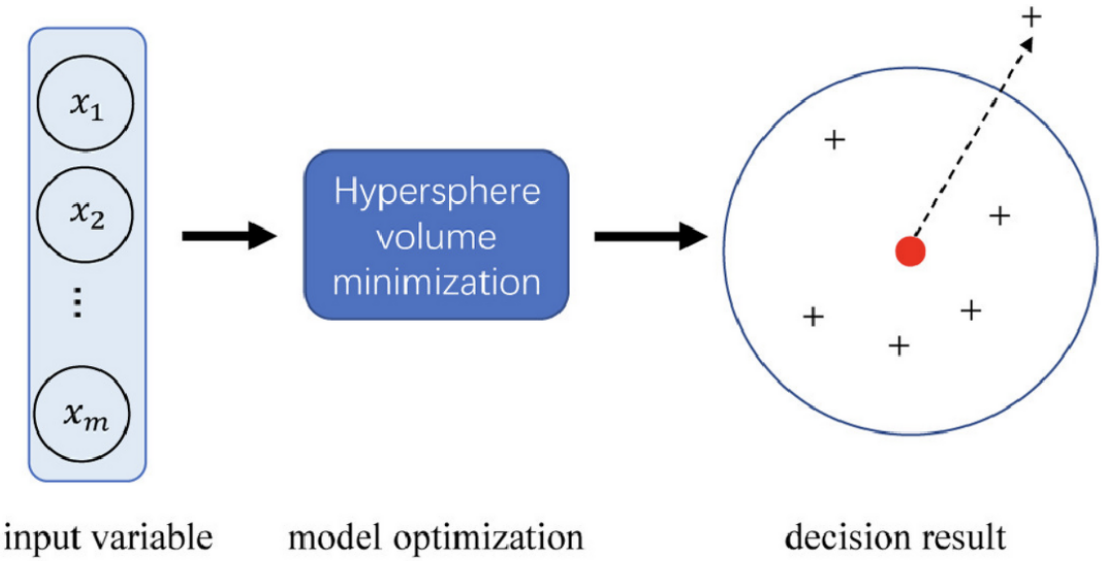
SVM vs SVDD

- 

- 

- SVM : SVDD의 근간을 이루는 SVM은 서로 다른 Class의 Sample들을 잘 분류하는 Classifier를 찾는 것
    - 데이터의 차원이 2차원일때는 직선 Classifier, 3차원일때는 평면 Classifider, ..., d차원일때는 d-1차원의 Hyperplane Classifier를 찾는 것
- SVDD : **Non-linear SVM**을 응용한 **One-class Classification**을 위한 대표적인 방법으로 이상치를 분류하는 기법
    - One-class Classification : Class-imbalance가 심한 경우 정상 sample만 이용하여 데이터 분류
    - = Semi-supervised Learning

SVDD 개념



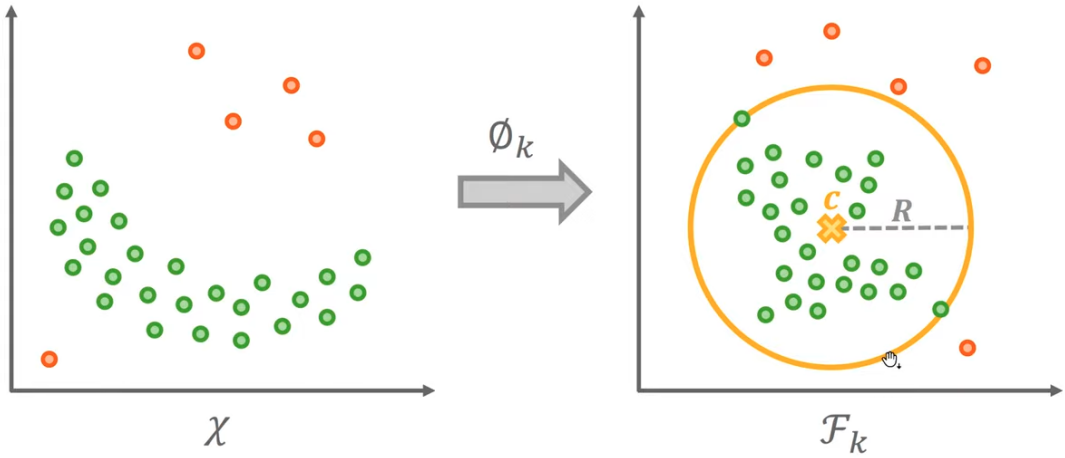
- Feature Space에서 정상 데이터를 둘러싸고 있는 Hypersphere를 찾고, 해당 **Hypersphere Boundary**를 통해 정상과 비정상 데이터를 구분함, 이 때 **Hypersphere를 찾을 때는 반지름을 최소화 하면서 많은 정상 데이터를 포함하는 것을 목표로 함**

SVDD의 Objective Function

- Feature Space에서 대부분의 정상 데이터를 둘러싸는 중심이 c인 가장 작은 Sphere의 경계를 찾음

SVDD 개요

- SVDD는 feature space에서 정상 데이터를 둘러싸는 가장 작은 구를 찾고, 해당 경계면을 기반으로 이상치를 탐지함
- Notation
  - ✓ χ: input space, $\mathcal{F}_k$: feature space, $\emptyset_k$: kernel function, R: radius, c: center



- D 차원 입력 공간이 n개의 데이터로 구성되어 있을 때, 중심이 c이고, 반경이 R인 Sphere를 이용
- 이후 각 학습 데이터 X_i를 Kernel Function으로 맵핑 시킨 $\phi_k(x_i)$와 중심 c 사이의 거리가 반지름인 R을 초과하는 경우 Panelty ξ 를 부과
- Kernel Function을 통해 Mapping된 Input의 Feature들과 Sphere 중심의 차이가 $R^2 + \xi_i$ 를 더한 것 보다 작다고 정의
- 따라서 아래 식을 최적화 하여 최소한의 반지름 R을 가지는 sphere을 찾음
- **즉, Left term은 가장 작은 Sphere 경계를 찾기 위함, Right term은 대부분의 정상 데이터를 둘러싸는 경계를 찾기 위함**

$$\min_{R,\boldsymbol{c},\boldsymbol{\xi}} \quad R^2 + \frac{1}{\nu n}\sum_i \xi_i$$

$$\text{s.t.} \quad \|\phi_k(\boldsymbol{x}_i) - \boldsymbol{c}\|^2_{\mathcal{F}_k} \le R^2 + \xi_i,\ \xi_i \ge 0, \quad \forall i.$$

-

- Parameters

- 비정상/정상을 잘 분류할 수 있는 가장 작은 Sphere를 찾기위해 사용

- R : radius
- c : Center Point
- ξ : Penalty Term for soft margin

- Hyper-parameters

- 비정상/정상을 잘 분류할 수 있는 경계값을 찾는 파라미터

- ϕ_k : Kernel Function (기존 Feature들의 차원을 변환)
- v : Trade-off (volume of sphere and violations of boundary)

- v는 Sphere의 크기를 결정하며, 반지름 R과 오차 간의 Trade-off 관계에 있음

- v가 작아질수록 Sphere의 크기가 커져서 대부분의 정상 클래스가 해당 Sphere에 들어가게 되지만, 비정상 클래스가 혼합될 수 있음
- 반면 v가 커질수록 Sphere가 작아지게 되고 이로 인해 비정상 클래스를 잘 검출할 수 있지만 몇몇 정상 클래스가 Sphere 경계면 바깥으로 나갈 수 있음
- **v에 대한 설명 추가**
  - SVDD 논문 : https://homepage.tudelft.nl/a9p19/papers/ML_SVDD_04.pdf

    Analogous to the Support Vector Classifier (Vapnik, 1998) we define the error function to minimize:

    $$F(R, \mathbf{a}) = R^2 \tag{1}$$

    with the constraints:

    $$\|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2, \quad \forall i \tag{2}$$

    To allow the possibility of outliers in the training set, the distance from $\mathbf{x}_i$ to the center $\mathbf{a}$ should not be strictly smaller than R^2, but larger distances should be penalized. Therefore we introduce slack variables $\xi_i \geq 0$ and the minimization problem changes into:

    $$F(R, \mathbf{a}) = R^2 + C \sum_i \xi_i \tag{3}$$

    with constraints that almost all objects are within the sphere:

    $$\|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0 \quad \forall i \tag{4}$$

  - The parameter C controls the trade-off between the volume and the errors.
    - error function을 최소화 하기 위해 수식 2와 같이 정의하며,
      - $F(R, a) = R^2$
      - $||x_i - a||^2 \leq R^2$
    - 여기서 x_i 에서 center a (여기서는 center 가 a로 표기)의 거리는 R^2 보다 엄격하게 작지 않아야하지만, larger distance를 가질 경우 penalty가 부과되어야 함
    - 따라서 slack variable $\xi \geq 0$ 를 정의하여 아래와 같이 식을 만듦
      - $F(R, a) = R^2 + C \sum_i \xi_i$
    - **그러므로 여기서의 C는 ξ 를 적용하기 위한 cost function 이라고 볼 수 있으며, volume 및 error 사이의 trade-off를 조정 할 수 있는 파라미터임**

For hyperplane $\mathbf{w}$ which separates the data $\mathbf{x}_i$ from the origin with margin ρ, the following holds:

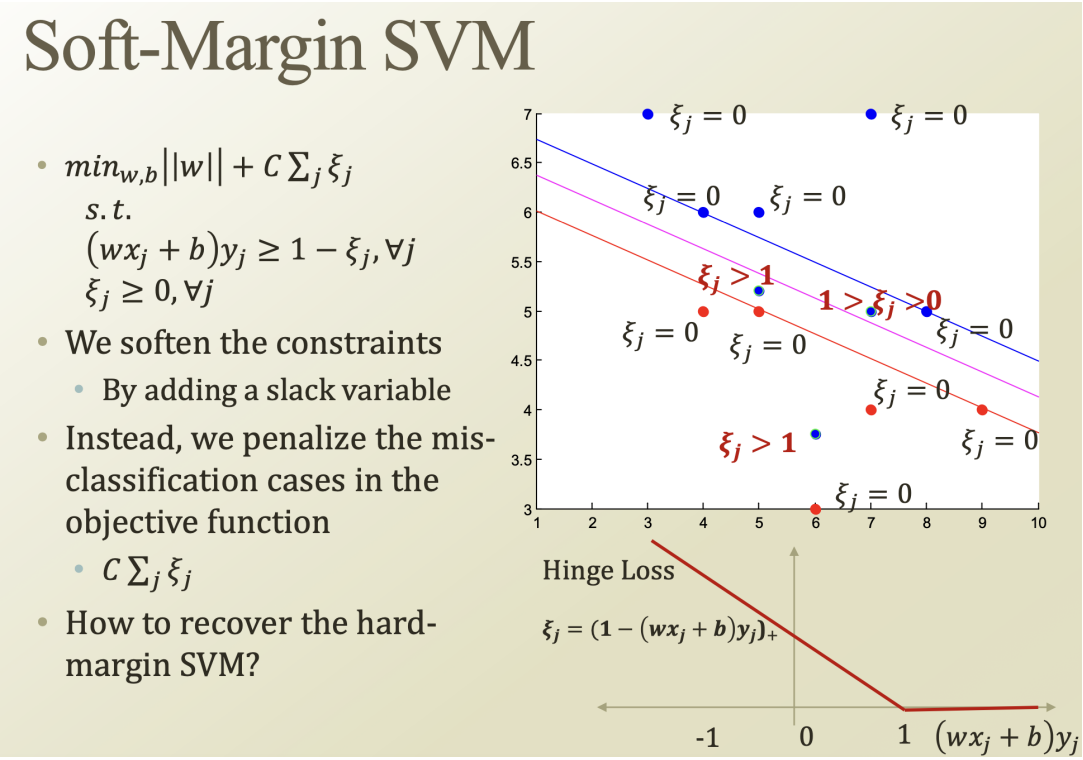
$$\mathbf{w} \cdot \mathbf{x}_i \geq \rho - \xi_i \quad \forall i \quad \xi_i \geq 0 \tag{16}$$

where ξ_i accounts for possible errors. Schölkopf minimizes the structural error of the hyperplane, measured by $\|\mathbf{w}\|$. This results in the following minimization problem:

$$\min_{\mathbf{w},\rho,\xi} \frac{1}{2}\|\mathbf{w}\|^2 - \rho + \frac{1}{\nu N}\sum_i \xi_i \tag{17}$$

with constraints (16). The regularization parameter $\nu \in (0, 1)$ is a user defined parameter indicating the fraction of the data that should be separated and can be compared with the parameter C in the SVDD. Here we will call this method the ν-SVC.

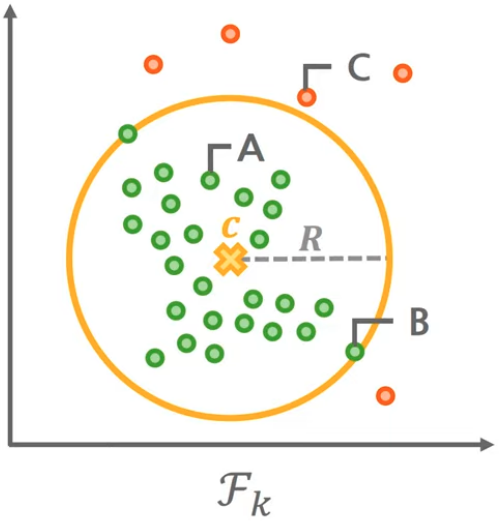
- 이에 따라 위 식 17에서 등장하는 $\frac{1}{\nu N}$ 은 위에서 설명한 C라고 볼 수 있으며, 여기서 ν 는 0에서 1의 값을 가질 수 있는 regularization parameter이고 이는 분리되어야 하는 데이터의 비율을 나타내는 사용자 정의 parameter에 해당
- 이번 SVM 강의 내용에서 Soft-Margin SVM 내용 설명에서 등장하는 C와 같은 역할임. 즉 panelty ξ 를 적용하기 위한 cost function이라고 보면 됨



- 따라서 위 식을 최적화 하여 얻은 Hypersphere의 경계면을 기준으로 비정상/정상을 분류

Decision Function

- 구 경계면에 대한 데이터 포함 여부를 기반으로 데이터의 normal/abnormal 여부를 도출하며, decision function은 아래와 같음
- Decision function (+1: normal & −1: abnormal)
  - ✓ $f(x_i) = sign(R^2 - \|\emptyset_k(x_i) - c\|^2_{\mathcal{F}_k})$



A (inside): $R > \|\emptyset_k(x_i) - c\|_{\mathcal{F}_k} \rightarrow f(A) = +1$

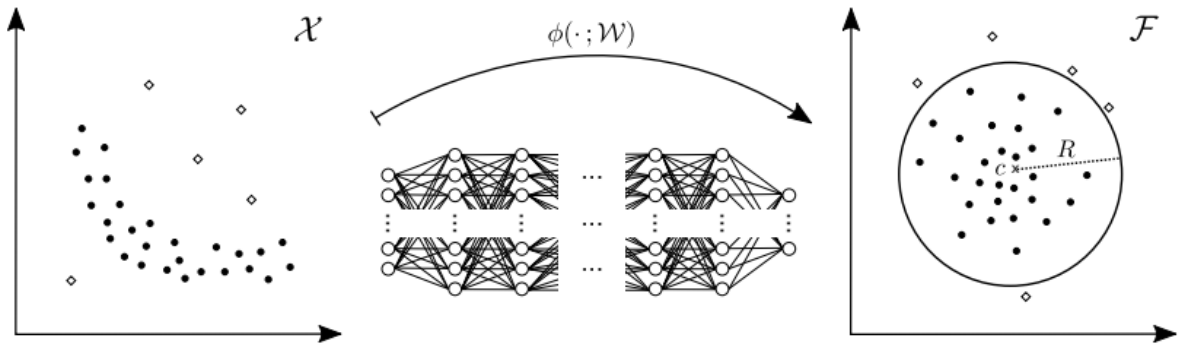
B (margin): $R = \|\emptyset_k(x_i) - c\|_{\mathcal{F}_k} \rightarrow f(B) = +1$

C (outside): $R < \|\emptyset_k(x_i) - c\|_{\mathcal{F}_k} \rightarrow f(C) = -1$

- SVDD의 단점
  - Kernel based model이기 때문에 데이터가 늘어날수록 연산량이 4 제곱 만큼 늘어나기 때문에 데이터가 많을 경우 연산의 비효율성 발생

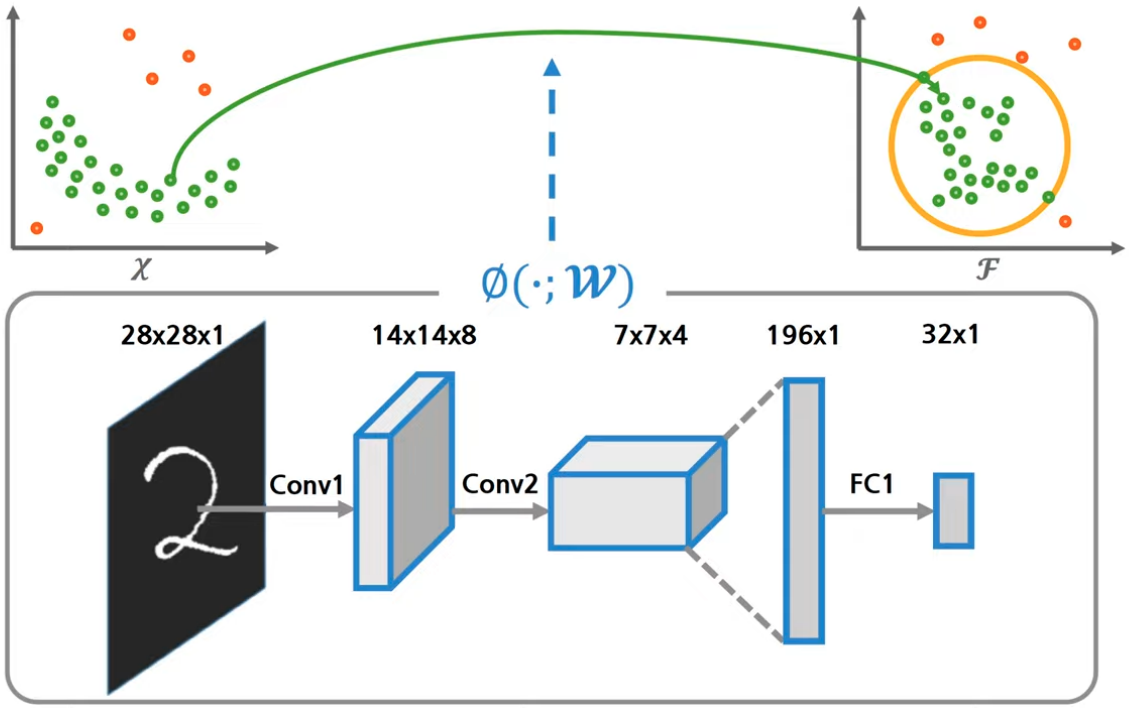
Deep SVDD

- Paper : Deep One-Class Classification (2018)
  - https://proceedings.mlr.press/v80/ruff18a.html



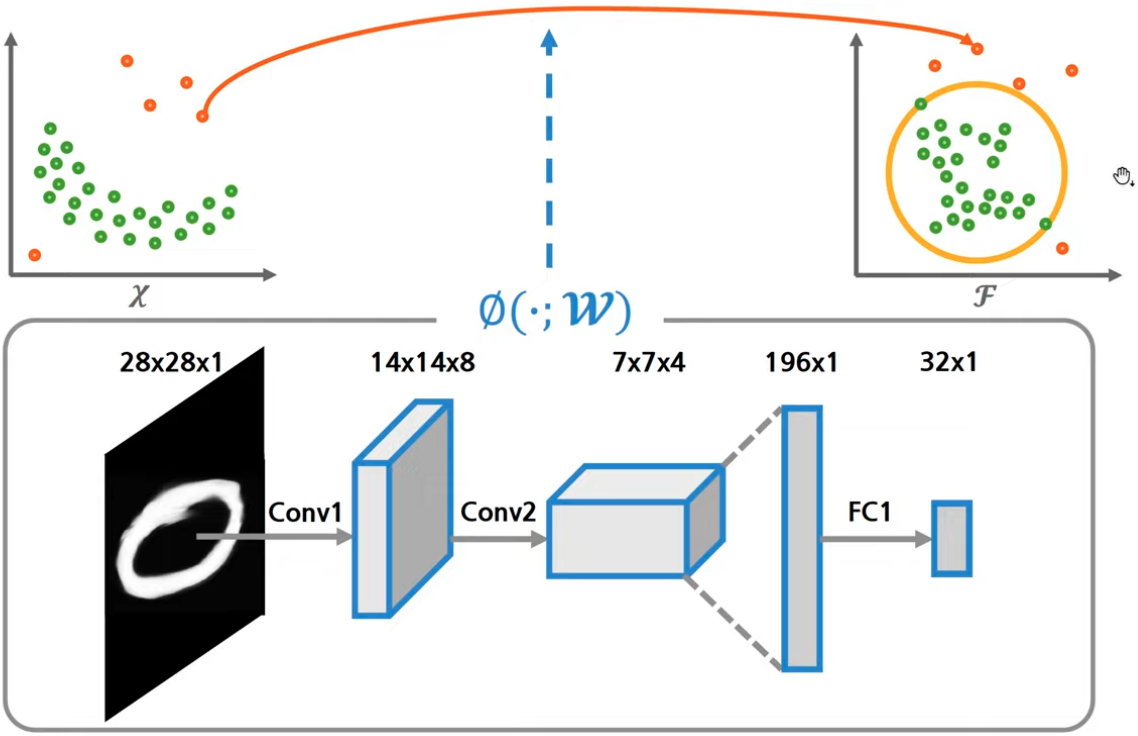
Deep SVDD 개요

- Normal example은 구 안으로 mapping 되도록 하는 $\mathcal{W}$와 R을 학습함



Deep SVDD 개요

- Abnormal example은 구 밖으로 mapping 되도록 하는 $\mathcal{W}$와 R을 학습함



- 기존 Kernel을 기반으로 한 SVDD와는 다르게 **딥러닝 기반으로 Feature Space 학습**
- Normal Data에 대해 Feature Space 상의 한 점에 모이도록 네트워크를 학습하는 방법
- 즉, Kernel Function을 딥러닝 방식으로 대체하여 기존 데이터 Point들을 매핑하는 가중치를 학습하는 것

Deep SVDD의 Objective Function

- Soft Boundary Deep SVDD

- Objective Function을 최적화하고, w는 정상 데이터의 공통되는 feature를 추출하여 각 데이터 point를 Sphere의 중심에 가깝게 매핑

$$\min_{R,W} R^2 + \frac{1}{vn}\sum_i \max\{0, \| \phi(x_i;W) - c \|^2 - R^2\} + \frac{\lambda}{2}\sum_l \| W^l \|_F^2$$

- c, v는 hyperparameter
- 첫번째 항 : Hypersphere의 volume을 최소화 하도록 유도
- 두번째 항 : Hypersphere의 반지름 R 보다 더 멀리 분포함 feature에 대한 panelty ξ 부여
- 세번째 항 : weight decay regularizer

- One-class Deep SVDD

- Soft Boundary Deep SVDD 방식에서 첫번째, 두번째 항을 합쳐 Center로부터 가깝게 모이도록 유도

$$\min_{\mathcal{W}} \quad \frac{1}{n}\sum_{i=1}^{n} \|\phi(\boldsymbol{x}_i;\mathcal{W}) - \boldsymbol{c}\|^2 + \frac{\lambda}{2}\sum_{\ell=1}^{L} \|\boldsymbol{W}^\ell\|_F^2.$$

- Sphere의 중심 c와 변환된 차원의 데이터 point 사이의 거리를 최소화
- Weight Decay Regularizer 항으로 특정 가중치가 비정상적으로 커지는 것을 방지
- 각 데이터 Point들이 Sphere의 중심 c에 가깝게 매핑되도록 학습

- Anomaly Score
  - 학습이 완료된 뒤, test dataset에 대한 anomaly score를 아래와 같이 정의
  - $$s(x) = \| \phi(x;W) - c \|^2$$
  - ϕ 함수와 c의 거리가 anomaly score로 설정되어 정상, 비정상 데이터 판단
  - 즉, 신경망의 Output Feature에 대해 Hypersphere의 Center에서 멀리 떨어진 정도를 anomaly score로 사용

- MNIST, CIFAR-10에서의 실험 결과

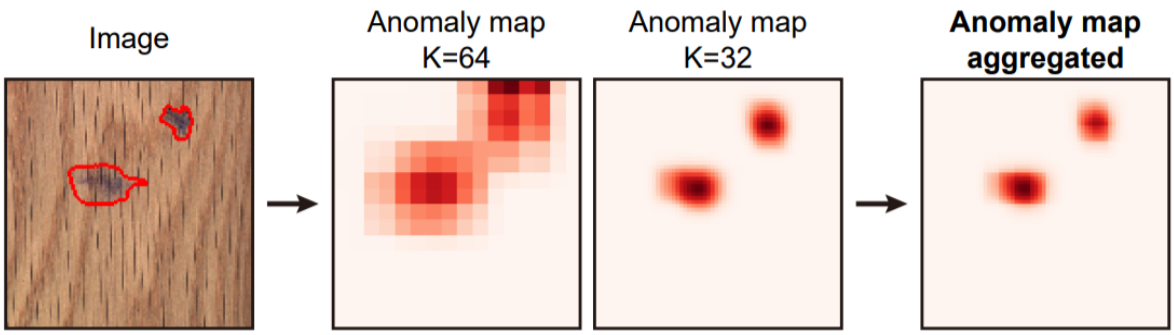
Table 1. Average AUCs in % with StdDevs (over 10 seeds) per method and one-class experiment on MNIST and CIFAR-10.

NORMAL CLASS	OC-SVM/ SVDD	KDE	IF	DCAE	ANOGAN	SOFT-BOUND. DEEP SVDD	ONE-CLASS DEEP SVDD
0	**98.6**±0.0	97.1±0.0	98.0±0.3	97.6±0.7	96.6±1.3	97.8±0.7	98.0±0.7
1	99.5±0.0	98.9±0.0	97.3±0.4	98.3±0.6	99.2±0.6	99.6±0.1	**99.7**±0.1
2	82.5±0.1	79.0±0.0	88.6±0.5	85.4±2.4	85.0±2.9	89.5±1.2	**91.7**±0.8
3	88.1±0.0	86.2±0.0	89.9±0.4	86.7±0.9	88.7±2.1	90.3±2.1	**91.9**±1.5
4	**94.9**±0.0	87.9±0.0	92.7±0.6	86.5±2.0	89.4±1.3	93.8±1.5	**94.9**±0.8
5	77.1±0.0	73.8±0.0	85.5±0.8	78.2±2.7	88.3±2.9	85.8±2.5	**88.5**±0.9
6	96.5±0.0	87.6±0.0	95.6±0.3	94.6±0.5	94.7±2.7	98.0±0.4	**98.3**±0.5
7	93.7±0.0	91.4±0.0	92.0±0.4	92.3±1.0	93.5±1.8	92.7±1.4	**94.6**±0.9
8	88.9±0.0	79.2±0.0	89.9±0.4	86.5±1.6	84.9±2.1	92.9±1.4	**93.9**±1.6
9	93.1±0.0	88.2±0.0	93.5±0.3	90.4±1.8	92.4±1.1	94.9±0.6	**96.5**±0.3
AIRPLANE	61.6±0.9	61.2±0.0	60.1±0.7	59.1±5.1	**67.1**±2.5	61.7±4.2	61.7±4.1
AUTOMOBILE	63.8±0.6	64.0±0.0	50.8±0.6	57.4±2.9	54.7±3.4	64.8±1.4	**65.9**±2.1
BIRD	50.0±0.5	50.1±0.0	49.2±0.4	48.9±2.4	**52.9**±3.0	49.5±1.4	50.8±0.8
CAT	55.9±1.3	56.4±0.0	55.1±0.4	58.4±1.2	54.5±1.9	56.0±1.1	**59.1**±1.4
DEER	66.0±0.7	**66.2**±0.0	49.8±0.4	54.0±1.3	65.1±3.2	59.1±1.1	60.9±1.1
DOG	62.4±0.8	62.4±0.0	58.5±0.4	62.2±1.8	60.3±2.6	62.1±2.4	**65.7**±2.5
FROG	74.7±0.3	**74.9**±0.0	42.9±0.6	51.2±5.2	58.5±1.4	67.8±2.4	67.7±2.6
HORSE	62.6±0.6	62.6±0.0	55.1±0.7	58.6±2.9	62.5±0.8	65.2±1.0	**67.3**±0.9
SHIP	74.9±0.4	75.1±0.0	74.2±0.6	**76.8**±1.4	75.8±4.1	75.6±1.7	75.9±1.2
TRUCK	75.9±0.3	**76.0**±0.0	58.9±0.7	67.3±3.0	66.5±2.8	71.0±1.1	73.1±1.2

- Deep SVDD의 단점?
    - 하나의 Center 기준으로 데이터들을 Sphere에 가깝게 매핑하기 때문에 비정상 데이터와 정상 데이터 간의 차이가 매우 작을 때 Deep SVDD를 적용한다면 비정상 데이터를 제대로 추출 못할 수도 있음

Further ...

- Deep SVDD를 최적화 할 수 있는 방법
    - https://hongl.tistory.com/78
- multiscale vector data description (MVDD) (2020)
    - Timeseries anomaly detection 분야에서 Deep SVDD와 Dilated RNN을 이용하여 Dilated RNN을 기반으로 short/long-term 정보를 포함함 multiscale temporal features를 추출한 뒤, differentiable hierarchical clustering을 통해 multiscale temporal features를 통합하고 이를 기반으로 이상치를 탐지
    - http://dsba.korea.ac.kr/seminar/?mod=document&uid=1388
- Patch SVDD (2020)



    - SVDD를 Patch-based 방법으로 바꾼 것

    - Deep SVDD가 딥러닝을 이용하여 data-dependent한 representation을 추출하였다면, Patch SVDD는 이를 Patch-wise 방식으로 바꾼 것이고, patch들의 높은 분산과 self-supervised 학습 방식을 사용함
    - Multi-scale로 검출 결과를 추출한 다음 이를 결합하여 이상치를 탐지
    - 관련 자료
        - https://ysco.tistory.com/17
        - https://openaccess.thecvf.com/content/ACCV2020/papers/Yi_Patch_SVDD_Patch-level_SVDD_for_Anomaly_Detection_and_Segmentation_ACCV_2020_paper.pdf

참고자료

- Anomaly Detection
    - https://hoya012.github.io/blog/anomaly-detection-overview-1/
    - https://ffighting.tistory.com/entry/%EB%94%A5%EB%9F%AC%EB%8B%9D-Anomaly-Detection
- SVDD
    - https://wsshin.tistory.com/3
    - https://hoya012.github.io/blog/anomaly-detection-overview-1/
    - https://wikidocs.net/3155
    - https://yupsung.blogspot.com/2021/02/one-class-svm-svdd.html
    - http://dsba.korea.ac.kr/seminar/?mod=document&uid=1327
- Deep SVDD
    - https://ys-cs17.tistory.com/50
    - https://ffighting.tistory.com/entry/OC-Deep-SVDD-%ED%95%B5%EC%8B%AC-%EB%A6%AC%EB%B7%B0
    - http://dsba.korea.ac.kr/seminar/?mod=document&uid=1327