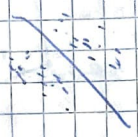
Lecture Kernel Machines 3

Anomaly Detection

- security (credit cards)
- experimental outliers (EEG)

Problem: How to define model?

Two class: easy:



One class hard

Density Based

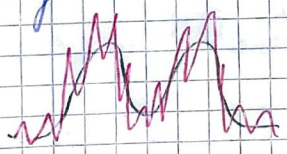
Kernel Density Estimation

normalization

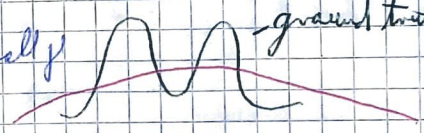
Define probability $p(x) = \frac{1}{N} \sum_{i=1}^{N} k(x, x_i)$

often: Gaussian kernel: $k(x, x_i) = e^{-\gamma \|x - x_i\|^2}$

large γ



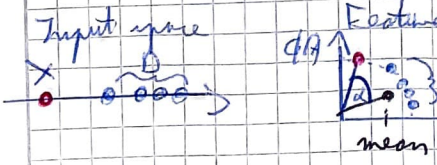
small γ



we don't have to learn, just define γ

If k induces a feature map:

$$p(x) = \frac{1}{N} \sum_{i=1}^{N} k(x, x_i) = \frac{1}{N} \sum_{i=1}^{N} \langle \phi(x), \phi(x_i) \rangle = \langle \phi(x), \underbrace{\frac{1}{N} \sum_{i=1}^{N} \phi(x_i)}_{\text{mean}} \rangle$$



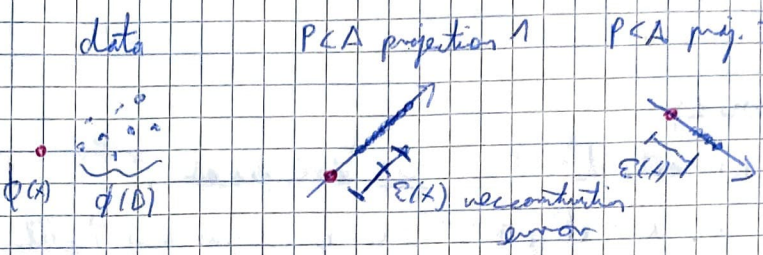
Outlier score: $o(x) = -\log p(x)$

Reconstruction based

Kernel PCA

Learn a reconstruction model of the data $x \mapsto \text{proj}_p(x)$ and detect outlier through large reconstruction error.

$$O(x) = \|\phi(x) - \sum_{i=1}^{q} u_i u_i^T \phi(x)\|^2 = k(x,x) - \sum_{i=1}^{q} (u_i^T \phi(x))^2$$



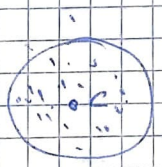
Boundary based

SVDD, OC-SVM

Learn a separating surface between the in- and outlier data (e.g., sphere)

Support Vector Data Description: sphere with center c in feature space and radius R

- R can be optimized with constrained quad. prog.



Primal: $\min_{R,c,\xi} R^2 + \frac{1}{N\nu} \sum_{i=1}^{N} \xi_i$

$$\text{s.t. } \forall_{i=1}^{N}: \|\phi(x_i) - c\|^2 \leq R^2 + \xi_i \geq 0, \quad \xi_i \geq 0$$

ξ_i: slack variables (like soft-margin SVM)

N: data points

ν: hyperparameter

Pro:

- unsupervised
- optimization convex
- ν: upper bound of fraction of outlier

Dual:

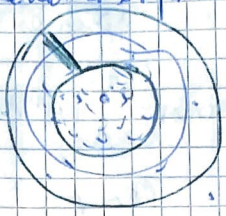
$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i k(x_i,x_i) - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j k(x_i,x_j)$$

$$\text{s.t. } \sum_{i=1}^{N} \alpha_i = 1, \quad \forall_{i=1}^{N}: 0 \leq \alpha_i \leq \frac{1}{N\nu}$$

Prefer Dual over primal if $\phi(x)$ continuous or $d \gg N$

Semi-Supervised Anomaly Detection

- enable experts to verify uncertain queries



Summary

- anomaly detection not classification
  → usually treated as unsupervised task