School of Computing and Information Systems (CIS)
The University of Melbourne
COMP90073
Security Analytics
Tutorial exercises: Week 7

1. Give example of 2 applications that it is better to use adaptive window over sliding window in data stream anomaly detection. Justify your answer.

2. We used the following example to explain the step by step iLOF's measurements update. We included point 11 in *reachdist* update (Figure 1) but not in lrd update (Figure 2). Explain why, given *k*=2.

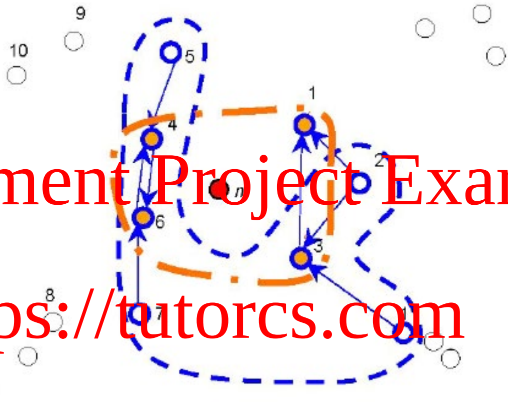

Figure 1: *reachdist* update

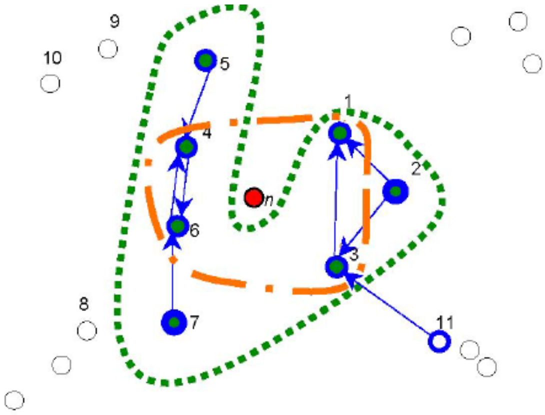

Figure 2: lrd update

Solution: We update lrd value of point *p* if

- The *k*-neighbourhood of the point *p* changes,
- *Reachdist* from point *p* to one of its *k*-neighbours changes.

3. In iLOF deleting a point p_i from the existing dataset *always* increases the k-distances of R_k-NN of p_i. Justify the reason.

   See lecture notes <Week6-L1.pdf>.

4. In what case performance of MiLOF resembles to iLOF?

   Solution: As the width of the summarization bucket/window decreases, MiLOF begins to resemble iLOF, and in the limit (when there is no historical retention by summarization), MiLOF reduces to iLOF.

5. In the lecture we saw how we can derive SVDD's dual formulation from its primal formulation. Now given OCSVM's primal formulation as below, derive its dual formulation.

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

$$\text{s.t.}$$

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

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

Solution:

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

- $\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n}\alpha_i\,\phi(x_i) = 0$ $\qquad w = \sum_{i=1}^{n}\alpha_i\,\phi(x_i)$
- $\frac{\partial L}{\partial \rho} = -1 - \sum_{i=1}^{n}\alpha_i = 0$ $\qquad \sum_{i=1}^{n}\alpha_i = -1$
- $\frac{\partial L}{\partial \xi_i} = \frac{1}{\nu n} - \alpha_i - \gamma_i = 0$ $\qquad \frac{1}{\nu n} = \alpha_i + \gamma_i$

$$\begin{aligned}
L(w,\rho,\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\gamma}) &= \frac{1}{2}w^T w + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho - \sum_{i=1}^{n}\alpha_i w^T\phi(x_i) - \rho\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}\alpha_i\,\xi_i - \sum_{i=1}^{n}\gamma_i\xi_i \\
&= \frac{1}{2}w^T w + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho - \sum_{i=1}^{n}\alpha_i w^T\phi(x_i) - \rho\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}\alpha_i\,(\xi_i + \gamma_i) \\
&= \frac{1}{2}w^T w + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho - \sum_{i=1}^{n}\alpha_i w^T\phi(x_i) - \rho\sum_{i=1}^{n}\alpha_i - \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i \\
&= \frac{1}{2}w^T w - \rho - \sum_{i=1}^{n}\alpha_i w^T\phi(x_i) - \rho\sum_{i=1}^{n}\alpha_i \\
&= \frac{1}{2}w^T w - \rho - w^T w - \rho\sum_{i=1}^{n}\alpha_i \\
&= -\frac{1}{2}w^T w - \rho - \rho\sum_{i=1}^{n}\alpha_i
\end{aligned}$$

$$= -\frac{1}{2}w^T w - \rho + \rho$$
$$= -\frac{1}{2}w^T w$$

$$\underset{\alpha}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \, \alpha_j \, k(x_i, x_j)$$
$$\text{s.t.} \qquad 0 \le \alpha_i \le \frac{1}{\nu n}, \quad \sum_{i=1}^{n} \alpha_i = 1$$

6. Use OneClassSVM in Splunk to perform unsupervised outlier detection. Some useful information regarding the parameters: https://scikit-learn.org/stable/modules/generated/sklearn.svm.OneClassSVM.html

7. You may use LIBSVM (https://www.csie.ntu.edu.tw/~cjlin/libsvm/) for the following exercises. The web page provides the necessary information for parameter tuning.
   Download the KDDCUP data set from the UCI Machine Learning Repository https://archive.ics.uci.edu/ml/datasets/kdd+cup+1999+data
   a. Use SVDD and OCSVM to identify the attacks.
   b. How many data points are common among the identified anomalies using different methods?