Student Number: □

The University of Melbourne
Semester 2 Assessment 2022

School of Computing and Information Systems

COMP90073 Security Analytics

Reading Time: 15 minutes.
Writing Tim
This paper
Common Co

Authorised material.

Instructions

- This pa 60 marks in total.
- There a
- Answer d then upload the complet re unable to print the exa may write on your own bla nswers.
- You ma need to revise an answer
- You mu sistance from anyone else aging, chat rooms, email, t anyone else taking the exam. You must not post answers to the questions or discussion of the questions online. Failure to comply with these instructions may be considered as academic misconduct.
- You are free to use the course materials and your laptop/PC in this exam but note that there is a 2-hour time window for the exam hence you should be mindful of the time spent using such resources.
- Answer the questions as clearly and precisely as you can.
- Your writing should be clear. Unreadable answers will be deemed wrong. Excessively long answers or irrelevant information may be penalised.
- For numerical methods, marks will be given for applying the correct method.

Library: This paper may not be reproduced or held by the Baillieu Library.

Section A: Short Answer Questions (Use your own words to provide a short explanation to each question) **[10 marks in total]**

1. What is the pattern for a typical DNS amplification attack, and why? [1 marks]

Answer:

2. What type s? [1 marks]

   (a) Accou

   (b) Source

   (c) Applic

   (d) Wheth e

Answer:

3. Output of a bel. Provide an example [1 marks]

Answer:

4. Supervised machine learning models are not used for anomaly detection because in anomaly detection problems the data is highly imbalanced. One common solution to mitigate imbalance training is to over-sample the positive class (or under-sample the negative class). Are such solutions effective for anomaly detection problems? Justify your answer. [1 marks]

Answer:

5. Most of the anomaly detection methods introduced in this subject assume the training data is clean (i.e., not noisy), otherwise, their performance can significantly be impacted. Name two anomaly detection methods that are less susceptible to noisy data and discuss why they are more resilient. [1 marks]

Answer:

6. In Support Vector Data Description (SVDD) what are the training samples with $0 < \alpha$ [1 marks]

Answer:

7. Which of t nple for the given data p [1 marks]

Answer:

8. Adversarial training is an effective defence method against adversarial attacks. How does it augment the training dataset? [1 marks]

Answer:

9. One limitation of adversarial training is that it degrades the model's performance on clean data. Why is that? [1 marks]

Answer:

10. In adversarial attacks against reinforcement learning models, the attacker does not need to perturb every state observed by the agent. What is the heuristic method that decides whether to poison an observed state? [1 marks]

Answer:

Section B: Method and calculation Questions **[30 marks in total]**

11. You are a security expert working for MBank Financial Group. Your responsibility is to secure the company's IT systems, in particular, Payroll, Customer Relationship Management System and brochure hosting site.

    (a) How do you measure the confidentiality of the information you need to protect, and how can it be applied to information in those three systems? [2 marks]

    Answer:

    (b) What ? [1 marks]

    Answ

12. One recentl cial Group's online share rised access to a custom will cause a major impa and ratings of the explo

Metrics	Rating
Skill (High skill level required → low or no skill required)	2
Ease of Access (very difficult to do → very simple to do)	2
Incentive (high incentive → Low incentive)	5
Resource (requires expensive or rare equipment → no resources required)	3

(a) What is the likelihood score? [1 marks]

Answer:

(b) What is the risk level? [1 marks]

Answer:

(c) What is the recommended action, and why? Choose the appropriate answer, and briefly explain your choice.

i. Im ecide to not

pr

ii. Ac

iii. Ac

[1 marks]

Answ

13. The XLeag Intellectual Property is d the annualised rate

(a) What [1 marks]

Answer:

(b) What is the annualised loss expectancy? [1 marks]

Answer:

14. The table below shows a list of items, use FP-growth to identify frequent patterns with Min_sup=3. Your work should include FP-tree, Conditional pattern base, Conditional FP-tree, and Frequent patterns. [3 marks]

TID	List of items
T100	{a, b, c, k, l, m, s}
T200	{f, a, b, c, d, g, i, m, p}
T300	{a, b, d, h, j, k, w}
T400	{b, c, k, m, p}
T500	{a, f, c, e, l, p, k, n}
T600	{f, c, e, p, w}

Answer:

15. Local outlier factor (LOF) is one of the most effective anomaly detection techniques, however, it struggles to identify group anomalies which can appear frequently in cyber security problems. How would you extend LOF to be able to detect group anomalies as well as point anomalies? Discuss how your solution achieves this goal. [1 marks]

Answer:

16. Which of tl ation forest (iForest) an [2 marks]

(a) iForest space from the mi

(b) iForest nomaly.

(c) iForest ent lengths, while i

(d) iForest can.

Answer:

17. In your own word explain how graph convolutional networks (GCNs) adapt the idea of convolution to graph networks and why such a solution is needed.

[1 marks]

Answer:

18. One class support vector machine (OCSVM) solves the following quadratic problem to generate the decision boundary,

$$\min_{w,\xi_i,\rho} \frac{1}{2}||w||^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho$$
$$s.t.$$
$$(w \cdot \phi(x_i)) \geq \rho - \xi_i, \forall i = 1, \cdots, n$$
$$\xi_i \geq 0, \forall i = 1, \cdots, n$$

What are t
we want to
maximise t
[2 marks]

Answer:

19. In Task 1 o
detection algorithm on extracted features from network traffic, and gave you a training, a test, and a validation set. To address this task, one of your classmates, Flora, takes the following steps:

(a) Flora starts by fitting a PCA (n_component = 20) on the validation set with all the 15 features (including stream ID, without label), and calls the output model as “PCA_{fitted}”.

(b) Then, Flora applies the PCA to the validation set, and denotes the reduced dataset (processed by PCA) as “$Data_{val\_PCA}$”.

(c) Afterwards, Flora trains DBSCAN on $Data_{val\_PCA}$, and fine-tunes the parameters to get the highest accuracy.

(d) Finally, Flora extracts features from the training and test datasets by applying the PCA_{fitted} model, and applies DBSCAN to both data sets.

Flora finds the False Positive (FP) rate is too high for the trained DBSCAN model. Can you give some suggestions on how effectively Flora can reduce the FP rate? Will your method affect its True Positive (TP) rate? [2 marks]

Answer:

20. A binary li [illegible] ssifies input x using the [illegible] is classified into the pos [illegible] ve class. As demonstrat [illegible] l sample x' against f fo [illegible] orthogonal to the decis [illegible]

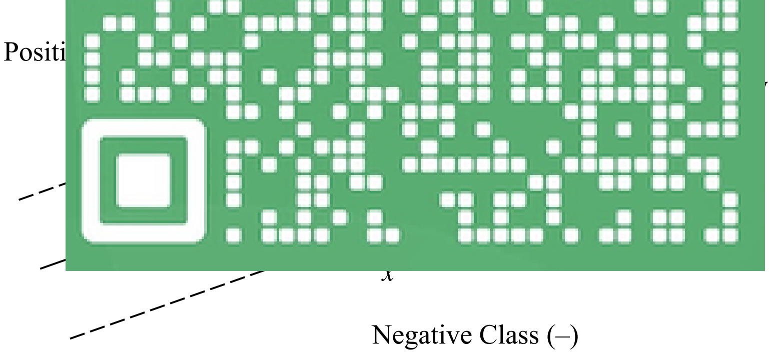

Figure 1: Generating an adversarial sample against a binary linear SVM classifier by moving the original input in a direction orthogonal to the decision boundary.

Suppose that $w = [4 \;\; 3]$, $b = 2$, and $x = [x_1 \;\; x_2]^T$, i.e., the input x is two dimensional. Generate an adversarial sample x' for point $(-1, 3)$ with the following two approaches:

(a) Fast gradient sign method (FGSM).