

Assignment -2

Q1: Given the values

$r_A = r_B = r_C = 0.5$  where  $r$  is the radius of the class enclosure

Initial adversary sample  $(x_0, y_0) = (0.1, -0.9)$

$$\text{Euclidian distance } A = \sqrt{(x^* - x_0)^2 + (y^* - y_0)^2}$$

class enclosures and their centroids are

green class  $(x_A, y_A) = (-0.8, 0.8)$

Blue class  $(x_B, y_B) = (0.05, -0.85)$

Red class  $(x_R, y_R) = (0.8, 0.8)$

① i) what would be class label assigned to the initial adversary sample

Ans: ① Centroid for green class  $(x_A, y_A) = (-0.8, 0.8)$

Initial adversary Sample  $(x_0, y_0) = (0.1, -0.9)$

Decision boundary of green class

$$(0.1 - (-0.8))^2 + (-0.9 - 0.8)^2 \leq r_A^2 \quad (r_A = 0.5)$$

$$\Rightarrow (0.1 + 0.8)^2 + (-1.7)^2 \leq 0.5^2$$

$$\Rightarrow (0.9)^2 + (1.7)^2 \leq 0.5^2$$

$$\Rightarrow 0.81 + 2.89 \leq 0.25$$

$$3.7 \leq 2.5 \quad (\text{NOT TRUE})$$

$\therefore$  The sample does not belong to the green class.

② Centroid for Red class  $(x_R, y_R) = (0.8, 0.8)$

Initial adversary Sample  $(x_0, y_0) = (0.1, -0.9)$

Decision boundary for Red class.

$$\Rightarrow (0.1 - 0.8)^2 + (-0.9 - 0.8)^2 \leq r_R^2 \quad (r_R = 0.5)$$

$$\Rightarrow (-0.7)^2 + (-1.7)^2 \leq 0.5^2$$

$$\Rightarrow 0.49 + 0.89 \leq 0.25$$

$$\Rightarrow 3.38 \leq 0.25 \text{ (NOT TRUE)}$$

$\therefore$  The sample does not belong to Red class.

③ Centroid for Blue class  $(x_B, y_B) = (0.05, -0.85)$

Initial adversary sample  $(x_0, y_0) = (0.1, -0.9)$

Decision boundary for Blue class

$$\Rightarrow (0.1 - 0.05)^2 + (-0.9 - (-0.85))^2 \leq \lambda_B^2 \quad (\lambda_B = 0.5)$$

$$\Rightarrow 0.0025 + (-0.05)^2 \leq 0.25$$

$$\Rightarrow 0.0025 + 0.0025 \leq 0.25$$

$$\Rightarrow 0.05 \leq 0.25 \text{ (TRUE)}$$

$\therefore$  The initial adversary sample belongs to  
Blue class

ii) Given the value.

$$(x^*, y^*) = (0.5, -0.1) \quad \& \quad (x_0, y_0) = (0.1, -0.9)$$

What is the strength of perturbation and what does the class label assigned to the final adversary sample?

$$\text{Strength of perturbation } A = \sqrt{(x^* - x_0)^2 + (y^* - y_0)^2}$$

$$\Rightarrow A = \sqrt{(0.5 - 0.1)^2 + (-0.1 - (-0.9))^2}$$

$$= \sqrt{(0.4)^2 + (0.8)^2}$$

$$= \sqrt{0.16 + 0.64}$$

$$= \sqrt{0.8} = 0.894$$

$\therefore$  Strength of perturbation

$$\boxed{A = 0.894}$$

① For green class

$$\text{centroid of green class } (x_G, y_G) = (-0.8, 0.8)$$

$$(x^*, y^*) = (0.5, -0.1)$$

Decision boundary for green class label

$$\Rightarrow (0.5 - (-0.8))^2 + (-0.1 - 0.8)^2 \leq \mu_G^2 \quad (\mu_G = 0.5)$$

$$\Rightarrow 1.3^2 + 0.9^2 \leq 0.5^2$$

$$\Rightarrow 1.69 + 0.81 \leq 0.25$$

$$\Rightarrow 2.5 \leq 0.25 \text{ (NOT TRUE)}$$

∴ The sample does not belong to the green class.

② For Red class

$$\text{centroid of Red class } (x_R, y_R) = (0.8, 0.8)$$

$$(x^*, y^*) = (0.5, -0.1)$$

Decision boundary for Red class label

$$\Rightarrow (0.5 - 0.8)^2 + (-0.1 - 0.8)^2 \leq \mu_R^2 \quad (\mu_R = 0.5)$$

$$\Rightarrow 0.3^2 + 0.9^2 \leq 0.5^2$$

$$\Rightarrow 0.09 + 0.81 \leq 0.25$$

$$\Rightarrow 0.9 \leq 0.25 \text{ (NOT TRUE)}$$

∴ The sample does not belong to the Red class.

③ For Blue class

$$\text{centroid of Blue class } (x_B, y_B) = (0.05, -0.85)$$

$$(x^*, y^*) = (0.5, -0.1)$$

Decision boundary for the Blue class.

$$\Rightarrow (0.5 - 0.05)^2 + (-0.1 - (-0.85))^2 \leq \mu_B^2 \quad (\mu_B = 0.5)$$

$$\Rightarrow 0.45^2 + (0.75)^2 \leq 0.25$$

$$\Rightarrow 0.2025 + 0.5625 \leq 0.25$$

$$\Rightarrow 0.765 \leq 0.25 \text{ (NOT TRUE)}$$

∴ The sample also does not belong to the Blue class.

Since the decision boundary of the sample are NOT TRUE in all the cases and so the sample can be rejected.

iii) Given the values  $(x^*, y^*) = (0.5, 0.5)$

what is the strength of perturbation and what would be the class label assigned to the final adversarial sample?

$$\text{strength of perturbation } A = \sqrt{(x^* - x_0)^2 + (y^* - y_0)^2}$$

$$\Rightarrow A = \sqrt{(0.5 - 0.1)^2 + (0.5 - (-0.9))^2}$$

$$A = \sqrt{0.4^2 + (0.5 + 0.9)^2}$$

$$= \sqrt{0.16 + 1.4^2}$$

$$= \sqrt{0.16 + 1.96} = \sqrt{2.12}$$

$$= \underline{\underline{1.456}}$$

∴ Strength of  
perturbation  $\boxed{A = 1.456}$

① For green class

$$\text{Centroid of green class } (x_a, y_a) = (-0.8, 0.8)$$

$$(x^*, y^*) = (0.5, 0.5)$$

Decision boundary of green class

$$\Rightarrow (0.5 - (-0.8))^2 + (0.5 - 0.8)^2 \leq u_a^2 \quad (\because u_a = 0.5)$$

$$\Rightarrow (0.5 + 0.8)^2 + (-0.3)^2 \leq (0.5)^2$$

$$\Rightarrow (1.3)^2 + 0.09 \leq 0.25$$

$$\Rightarrow 1.69 + 0.09 \leq 0.25$$

$$\Rightarrow 1.78 \leq 0.25 \text{ (NOT TRUE)}$$

∴ The sample does not belong to the green class.

② For Red class

$$\text{Centroid of red class } (x_c, y_c) = (0.8, 0.8)$$

$$(x^*, y^*) = (0.5, 0.5)$$

Decision boundary for the red class w.r.t Sample.

$$(0.5 - 0.8)^2 + (0.5 - 0.8)^2 \leq \alpha_r^2$$

$$(-0.3)^2 + (-0.3)^2 \leq 0.5^2$$

$$0.09 + 0.09 \leq 0.25$$

$$0.18 \leq 0.25 (\text{TRUE})$$

∴ The Adversary Sample belongs to the Red class

③ If  $\alpha_G = \alpha_R = \alpha_B = 0.5$ , what are the minimal and maximum strengths of perturbation A if this is an error generic Attack?

Ans: Given the values of radius.

$$\alpha_G = \alpha_R = \alpha_B = 0.5$$

W.E.T. Equation of line  $y = mx + c$

$$\text{where } m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$\text{Equation of circle: } (x - h)^2 + (y - k)^2 = \alpha^2$$

① For Red class

$$(x_2, y_2) = (0.8, 0.8)$$

$$(x_1, y_1) = (0.1, -0.9)$$

$$\therefore \text{slope } m = \frac{0.8 - 0.9}{0.8 - 0.1}$$

$$= \frac{1.7}{0.7} = 2.43$$

Equation of

$$\text{line } y = mx + c$$

$$0.8 = 2.43x + C$$

$$0.8 = 1.94 + C$$

$$\Rightarrow C = -1.14$$

Equation of circle  $(x-h)^2 + (y-k)^2 = r^2$   
 (Centroid of ued class  $(h,k) = (0.8, 0.8)$ )  
 $r = \text{radius of circle} = 0.5$

$$\Rightarrow (x-0.8)^2 + (y-0.8)^2 = 0.5^2 \quad (\text{where } y = 2.43x - 1.14)$$

$$\Rightarrow (x-0.8)^2 + (2.43x - 1.14 - 0.8)^2 = 0.5^2$$

$$\Rightarrow x^2 + 0.64 - 1.6x + (2.43x - 1.94)^2 = 0.25$$

$$\Rightarrow x^2 + 0.64 - 1.6x + 5.86x^2 - 9.38x + 3.76 = 0.25$$

$$\Rightarrow 6.86x^2 - 10.98x + 4.15 = 0$$

Solving the above equation for the value of  $x$  by using quadratic equations

Roots of quadratic equation  $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

$$x = \frac{(-(-10.98)) \pm \sqrt{(10.98)^2 - 4(4.15)(6.86)}}{2(6.86)}$$

$$= \frac{10.98 \pm \sqrt{120.56 - 113.88}}{13.72}$$

$$= \frac{10.98 \pm \sqrt{6.68}}{13.72}$$

$$= \frac{10.98 \pm 2.58}{13.72}$$

$$\therefore x_1 = 0.99$$

or

$$x_2 = 0.61$$

Now  $y_1 = 2.43 \times 0.99 - 1.14$

$$y_1 = 2.4 - 1.14$$

$$y_1 = 1.26$$

$$y_2 = 2.43 \times 0.61 - 1.14$$

$$y_2 = 1.48 - 1.14$$

$$y_2 = 0.34$$

Strength of perturbation

$$\textcircled{1} (x_1, y_1) = (0.99, 1.26)$$

$$\text{Strength of perturbation } A = \sqrt{(0.61 - 0.1)^2 + (1.26 + 0.99)^2}$$

$$A = \sqrt{(1.09)^2 + (2.16)^2}$$

$$A = \sqrt{1.1881 + 4.6656}$$

$$A = \sqrt{5.8537}$$

$$A = \underline{\underline{2.42}}$$

$$\textcircled{2} (x_2, y_2) = (0.61, 0.34)$$

$$\text{Strength of perturbation } A = \sqrt{(0.61 - 0.1)^2 + (0.34 + 0.9)^2}$$

$$A = \sqrt{(0.51)^2 + (1.24)^2}$$

$$A = \sqrt{0.2601 + 1.5376}$$

$$A = \sqrt{1.797}$$

$$A = \underline{\underline{1.34}}$$

\textcircled{3} For green class

$$(x_2, y_2) = (-0.8, 0.8)$$

$$(x_1, y_1) = (0.1, -0.9)$$

Equation of line  $y = mx + c$

$$\text{where } m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{0.8 + 0.9}{-0.8 - 0.1} = \frac{1.7}{-0.9} = \underline{\underline{-1.89}}$$

$$y = mx + c$$

$$0.8 = -1.89(-0.8) + c$$

$$0.8 = 1.512 + c$$

$$\therefore c = -0.712$$

$$\text{Equation of circle } (x-h)^2 + (y-k)^2 = r^2$$

$$\text{Centroid of green class } (h, k) = (-0.8, 0.8)$$

$$\text{Radius of circle (green) } r_a = 0.5$$

$$\Rightarrow (x - (-0.8))^2 + (-1.89x - 0.712 - 0.8)^2 = (0.5)^2$$

$$\Rightarrow (x + 0.8)^2 + (-1.89x - 1.512)^2 = 0.25$$

$$\Rightarrow x^2 + 1.6x + 0.64 + 3.57x^2 + 2.29 - 2(-1.89)(1.512) = 0.25$$

$$\Rightarrow 4.57x^2 + 7.32x + 2.93 = 0.25$$

$$\Rightarrow 4.57x^2 + 7.32 + 2.68 = 0$$

Solving using quadratic equations.

$$\text{Roots of quadratic equation} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$x = \frac{-7.32 \pm \sqrt{(7.32)^2 - 4(4.57)(2.68)}}{2(4.57)}$$

$$x = \frac{-7.32 \pm \sqrt{53.58 - 48.99}}{9.14}$$

$$x = \frac{-7.32 \pm \sqrt{4.59}}{9.14} = \frac{-7.32 \pm 2.14}{9.14}$$

$$x_1 = -1.035$$

or

$$x_2 = -0.567$$

The value of

$$y_1 = mx + c$$

$$= (-1.89)(-1.035) - 0.712$$

$$= 1.956 - 0.712$$

$$y_1 = \underline{\underline{1.244}}$$

$$y_2 = (-1.89)(-0.567) - 0.7$$

$$= 1.872 - 0.712$$

$$= \underline{\underline{0.36}}$$

Strength of perturbation.

$$\textcircled{1} (x_1, y_1) = (-1.035, 1.244)$$

$$\text{Strength of perturbation } A_3 = \sqrt{(-1.035 - 0.1)^2 + (1.244 + 0.9)^2}$$

$$A_3 = \sqrt{(1.135)^2 + (2.144)^2}$$

$$= \sqrt{1.288 + 4.597} = \sqrt{5.885} \\ = \underline{2.42}$$

$$\textcircled{2} (x_2, y_2) = (-0.567, 0.36)$$

$$\text{Strength of perturbation } A_4 = \sqrt{(-0.567 - 0.1)^2 + (0.36 + 0.9)^2}$$

$$A_4 = \sqrt{(0.667)^2 + (1.26)^2}$$

$$= \sqrt{0.45 + 1.59}$$

$$= \sqrt{2.04} = \underline{1.43}$$

: The values of strength of perturbation for the Error generic Attack are.

The maximum value of  $A = \underline{2.42}$

& the minimum value of  $A = \underline{1.43}$

③ If  $v_A = v_R = v_B = 0.5$ , what are the minimal & maximum strengths of perturbation of A if this is error specific attack targeting the R class. (Red class)

Ans: Given the values.

Centroid of class Red  $(x_R, y_R) = (0.8, 0.8)$

$(x_0, y_0) = (0.1, -0.9)$

Radius of class red  $\underline{\underline{d_R = 0.5}}$

For Error specific attack targeting the class R  
the strength of perturbation

$$A = \sqrt{(x_R - x_0)^2 + (y_R - y_0)^2}$$

$$= \sqrt{(0.8 - 0.1)^2 + (0.8 + 0.9)^2}$$

$$= \sqrt{0.44 + 2.89}$$

$$= \sqrt{3.38} = 1.838$$

Maximum perturbation  $A_{\max} = A + \lambda_R = 1.838 + 0.5 = \underline{\underline{2.338}}$

Minimum perturbation  $A_{\min} = A - \lambda_R = 1.838 - 0.5 = \underline{\underline{1.338}}$

The values of Maximum and minimum perturbation can be given by  $A_{\max} = 2.338$  &  $A_{\min} = 1.338$

~~Q~~ If  $y_A = y_B = 0.25$ , what are the minimal & maximal strengths of perturbation A if this is an error specific attack targeting the G class

① For Red class

The equation of line & calculations of quadratic expression can be brought from the Question 2  
Equation  $y = 2.43x - 1.14$  where  $c = -1.14$   
of line.  $m = 2.43$ .

Quadratic Expression.  $6.86x^2 - 10.98x + 4.4 = (0.25)^2$   
(since  $y_A = 0.25$ )

$$\Rightarrow 6.86x^2 - 10.98x + 4.4 = 0.0625$$

$$\Rightarrow 6.86x^2 - 10.98x + 4.338 = 0$$

Solving the above quadratic equation.

Roots of quadratic equation. 
$$= \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

$$\begin{aligned} &= \frac{-(-10.98) \pm \sqrt{(10.98)^2 - 4(6.86)(4.338)}}{13.72} \\ &= \frac{10.98 \pm \sqrt{120.56 - 119.03}}{13.72} \\ &= \frac{10.98 \pm \sqrt{1.53}}{13.72} \\ &= \frac{10.98 \pm 1.237}{13.72} \end{aligned}$$

$$\therefore x_1 = 0.89$$

$$\begin{aligned} y_1 &= 2.43x(0.89) - 1.14 \\ &= \underline{1.023} \end{aligned}$$

(or)

$$y_2 = 0.71$$

$$\begin{aligned} y_2 &= 2.43(0.71) - 1.14 \\ &= 1.725 - 1.14 \\ &= \underline{0.585} \end{aligned}$$

Strength of perturbation.

$$\textcircled{1} (x_1, y_1) = (0.89, 1.023)$$

$$\text{Strength of perturbation } A_{\text{g}} = \sqrt{(0.89 - 0.1)^2 + (1.023 + 0.9)^2}$$

$$A_{\text{g}} = \sqrt{(0.79)^2 + (1.923)^2}$$

$$A_{\text{g}} = \sqrt{0.624 + 3.698} = \sqrt{4.322}$$

$$A_{\text{g}} = 2.079$$

$$\textcircled{2} (x_2, y_2) = (0.71, 0.585)$$

$$\text{Strength of perturbation } A_{\text{g}} = \sqrt{(0.71 - 0.1)^2 + (0.585 + 0.9)^2}$$

$$A_{\text{g}} = \sqrt{(0.61)^2 + (1.485)^2}$$

$$A_{\text{g}} = \sqrt{0.3721 + 2.205}$$

$$= \sqrt{2.5771} = \underline{1.665}$$

Q) For Green class

The values of slope & y intercept for the equation of line & Quadratic expression can be obtained from Question '2'

Equation of line  $y = (-1.89)x + (-0.712)$

Quadratic expression  $4.57x^2 + 7.32x + 2.93 = 0.25^2$

$$\Rightarrow 4.57x^2 + 7.32x + 2.8675 = 0$$

Solve the quadratic equation by finding the roots.

$$\text{roots} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$= \frac{-7.32 \pm \sqrt{7.32^2 - 4 \times 4.57 \times 2.8675}}{2 \times 4.57}$$

$$= \frac{-7.32 \pm \sqrt{53.58 - 52.4179}}{9.14}$$

$$= \frac{-7.32 \pm \sqrt{1.1621}}{9.14}$$

$$\therefore x_1 = -0.92$$

(or)

$$x_2 = -0.68$$

$$y_1 = (-1.89)(-0.92) - 0.712$$

$$= 1.7388 - 0.712$$

$$= \underline{\underline{1.03}}$$

$$y_2 = (-1.89)(-0.68) - 0.712$$

$$= 1.2852 - 0.712$$

$$= \underline{\underline{0.54}}$$

$$\textcircled{1} (x_1, y_1) = (-0.92, 1.03)$$

Strength of perturbation  $A = \sqrt{(-0.92 - 0.1)^2 + (1.03 + 0.9)^2}$

$$A = \sqrt{(-1.02)^2 + (1.93)^2}$$

$$A = \sqrt{(1.04)^2 + 3^2}$$

$$A = \sqrt{4.76} = \underline{\underline{2.182}}$$

$$\textcircled{2} (x_1, y_1) = (-0.68, 0.57)$$

Strength of perturbation  $A = \sqrt{(-0.68 - 0.1)^2 + (0.57 + 0.9)^2}$

$$A = \sqrt{(-0.78)^2 + (1.07)^2}$$

$$A = \sqrt{0.61 + 2.161}$$

$$A = \sqrt{2.771}$$

$$A = \underline{\underline{1.605}}$$

$\therefore$  The values of strength of perturbation for the Error generic Attack are

the maximum value of  $A = \underline{\underline{2.182}}$

the Minimum value of  $A = \underline{\underline{1.605}}$

Q5 If  $v_A = v_B = v_g = 0.25$ , what are the minimal and maximal strengths of perturbation A if this is an error-specific attack targeting the A class (flipping the label to the green)

Ans Given the values

Centroid of the green class:  $(x_A, y_A) = (-0.8, 0.8)$

$(x_0, y_0) = (-0.1, 0.9)$

Radius of the class ball (in green)  $\lambda_g = 0.5$

For Error Specific attack targeting the class A  
the strength of perturbation

$$\begin{aligned} A &= \sqrt{(x_A - x_0)^2 + (y_A - y_0)^2} \\ &= \sqrt{(-0.8 - -0.1)^2 + (0.8 - -0.9)^2} \\ &= \sqrt{(-0.9)^2 + (1.7)^2} \\ &= \sqrt{0.81 + 2.89} = \sqrt{3.7} = \underline{\underline{1.924}} \end{aligned}$$

$$\begin{aligned} \text{Maximum perturbation } A_{\max} &= A + \lambda_g \\ &= 1.924 + 0.25 \\ &= \underline{\underline{2.174}} \end{aligned}$$

$$\begin{aligned} \text{minimum perturbation } A_{\min} &= A - \lambda_g \\ &= 1.924 - 0.25 \\ &= \underline{\underline{1.674}} \end{aligned}$$

The values of maximum & minimum perturbation can be given by

$$A_{\max} = \underline{\underline{2.174}}$$

$$A_{\min} = \underline{\underline{1.674}}$$

**INSE 6630 – Recent Developments in Information Systems Security**  
**Assignment 2. Due: 11:59 pm, Sunday, April 16, 2023.**

---

6. What observations can you draw from the results above?

**Answer**

1. The data distribution is random, but the decision border between the various classes is too narrow, which occasionally results in missing out the adversarial sample and some of the actual data points that are farther from the centre of the decision boundary of the given class.
2. The regions close to the class boundaries that are not classified into the various classes are referred to as the "grey zones." The rejection of points in these zones is caused by the decision borders' rigidity with respect to the circle's centre.
3. Changes in the decision boundaries of the classes do not cause a significant change in the strength of perturbation A.
4. The adversarial input of data affects the two distinct types of attack strategies created for carrying out evasion attacks, and modifications to it also affect the strength of perturbation.

2. Consider the ChatGPT that we introduced many in the course: <https://openai.com/blog/chatgpt>. Explore a potential use case of ChatGPT in the context of information system security and report what you have done and observed (copy the text or take a screenshot).

- a. Think of a specific security scenario or security problem where ChatGPT could be helpful and explain how it could be used to address that situation. First use one or two sentences to describe the use case in general (what task do you want ChatGPT to help you with), then specify (e.g., using a list of functionalities/logical conditions) the final result that you expect ChatGPT to provide to you at the end (codes, formulas, descriptions, comments, examples, etc.).
- b. Generate the prompts (a single question or a series of questions) that you will use to request the final and document these prompts.result
- c. Enter the prompts and document ChatGPT's answers to each of them.
- d. Choose the last prompt in the session and click "Regenerate response." Document the alternative answer it regenerates.
- e. Open a new session and try the same set of prompts. Document ChatGPT's new answers. Compare these new answers with the previous ones in 2c and highlight if there is any difference.
- f. Repeat 2b to generate and try a different set of prompts for the same use case. Document the new answers when you use a different set of prompts. Compare the difference, then comment on which set of prompts worked better, and why do you think so.

*A few tips for the prompt:*

- 1) Be specific: Be specific about the use case you're interested in. For example, you could ask for a use case of ChatGPT in detecting or preventing ransomware, or how to respond to it in specific operation scenarios. You may also ask for a use case in drafting rules for firewalls or

**INSE 6630 – Recent Developments in Information Systems Security**  
**Assignment 2. Due: 11:59 pm, Sunday, April 16, 2023.**

---

- spam filters, explain what deep packet inspection is, or suggest methods to detect anomalies in encrypted network traffic.
- 2) Provide context: Provide some context around the use case you're interested in. This could include information about the types of threats or attacks that are common or uncommon in the security landscape, or any specific challenges or limitations that currently exist in detecting and preventing such threats.
  - 3) Ask for details: In the prompt, ask for specific details on how ChatGPT could be used in the use case. For example, you could ask how ChatGPT could help identify new or emerging threats, detect at what stage an ongoing threat is, or how it could be used to analyze large amounts of data to detect patterns or anomalies. Consider these details and consider how many questions you will need as well as how they shall be ordered. For example, you may begin with a general one then specifies the details, or you may describe a procedure and give sequential questions on each step.
  - 4) Be open-ended: Finally, keep the prompt open-ended to allow for creative responses. Think outside the box and come up with innovative ways to apply ChatGPT to the security domain.

**Answer**

- a) **Use case:** Machine learning's role in an organization is to identify unusual network behavior and investigate the problem's underlying causes.

**Functionalities/Logical Conditions:**

- 1. Recognize a suspicious activity in a network.
- 2. Gathering datasets to offer analysis and guidance on how to investigate suspicious network activity.
- 3. Implementation the ML technique.

b) **Generated Prompts**

- 1. How to identify a suspicious activity in a network before the complete attack?
- 2. Can you provide the best ML technique used for this purpose and its implementation?

c) **Answers of ChatGPT for each prompt:**

*1. How to identify a suspicious activity in a network before the complete attack?*

To identify suspicious activity in a network before a complete attack, you can implement the following measures:

Implement network monitoring tools: Use network monitoring tools like intrusion detection systems (IDS) or intrusion prevention systems (IPS) to continuously monitor network traffic and identify any suspicious activity.

Set up anomaly detection: Configure anomaly detection tools to identify any unusual network behavior, such as a sudden surge in traffic, unusual network port usage, or unusual login attempts.

**INSE 6630 – Recent Developments in Information Systems Security**  
**Assignment 2. Due: 11:59 pm, Sunday, April 16, 2023.**

---

Establish a baseline: Establish a baseline for normal network behavior and use it to identify any deviation from it. This can help you detect suspicious activity early on.

Use threat intelligence: Stay up to date with the latest threat intelligence and use it to identify and block any known malicious IPs, domains, or URLs.

Educate employees: Educate employees on how to identify and report suspicious activity, such as unusual emails, pop-ups, or requests for sensitive information.

By implementing these measures, you can identify suspicious activity in a network before a complete attack occurs and take appropriate action to prevent it.

***2. Can you provide the best ML technique used for this purpose and its implementation?***

The best machine learning technique for detecting suspicious activity in a network will depend on the specific requirements of the system and the nature of the data. However, deep learning techniques have shown promising results in this area, particularly in detecting unknown and previously unseen types of attacks. One popular deep learning technique for network intrusion detection is a type of neural network called the Long Short-Term Memory (LSTM) network.

The LSTM network is a type of recurrent neural network that is designed to handle sequential data. It can be used to learn temporal dependencies in network traffic data and to identify suspicious patterns over time. Here is a high-level implementation of using LSTM networks for detecting suspicious network activity:

Data pre-processing: The first step is to pre-process the network traffic data, which may include cleaning the data, filtering out irrelevant features, and normalizing the data.

Feature extraction: Next, extract relevant features from the network traffic data. This can include things like the source and destination IP addresses, packet size, protocol type, and others.

Training: Use the pre-processed and feature-extracted data to train the LSTM model. The model should be trained on both normal and suspicious traffic data to help it learn to differentiate between the two.

Validation: Once the model is trained, validate it on a separate test dataset to evaluate its performance.

Deployment: Deploy the trained model to the production environment to continuously monitor network traffic and detect any suspicious activity.

Overall, implementing an LSTM network for detecting suspicious activity in a network involves pre-processing and feature extraction of the network traffic data, training the LSTM model, and deploying the model to the production environment.

**INSE 6630 – Recent Developments in Information Systems Security**  
**Assignment 2. Due: 11:59 pm, Sunday, April 16, 2023.**

---

**d) Alternative Answer for the last prompt:**

***Can you provide the best ML technique used for this purpose and its implementation?***

There is no one "best" machine learning technique for detecting suspicious activity in a network, as the most effective technique will depend on the specific requirements of the system, the nature of the data, and the resources available for implementation. However, here is an example of how an anomaly detection algorithm can be implemented for detecting suspicious activity in a network:

**Data Preparation:** First, a large dataset of network traffic records is collected, which includes both normal and suspicious activity. The data is then preprocessed to remove noise, outliers, and any irrelevant features.

**Feature Extraction:** Relevant features are extracted from the network traffic records, such as source and destination IP addresses, packet size, protocol type, etc. These features are then used to train the machine learning model.

**Model Training:** An unsupervised learning technique, such as a Gaussian Mixture Model (GMM), can be used to train the model. The GMM model builds a statistical model of normal network behavior and identifies any activity that deviates from that model as suspicious.

**Model Evaluation:** The performance of the model is evaluated on a separate dataset of network traffic records. The model's accuracy, precision, recall, and F1-score are calculated to determine its effectiveness in detecting suspicious activity.

**Deployment:** Once the model is trained and evaluated, it can be deployed to a production environment for real-time monitoring of network traffic. When suspicious activity is detected, an alert is triggered, and appropriate action can be taken.

It's important to note that this is just one example of how an anomaly detection algorithm can be implemented for detecting suspicious activity in a network. The specific implementation details will depend on the specific requirements of the system and the available resources.

**e) Answers from new session:**

***(Differences are highlighted in red color)***

**1. How to identify a suspicious activity in a network before the complete attack?**

There are several ways to identify suspicious activity in a network before a complete attack. Here are some tips:

**Monitor network traffic:** Regularly monitor network traffic for any anomalies or patterns that deviate from normal usage. This could include unusual spikes in traffic or unusual protocols being used.

**Analyze logs:** Analyze system logs to look for any suspicious behavior such as failed login attempts, changes to system settings, or other unusual activity.

**Use intrusion detection/prevention systems:** Implement an intrusion detection or prevention system that can analyze traffic and alert you to any suspicious activity. These systems can be configured to detect specific types of attacks or to look for unusual traffic patterns.

**INSE 6630 – Recent Developments in Information Systems Security**  
**Assignment 2. Due: 11:59 pm, Sunday, April 16, 2023.**

---

Keep software up to date: Keep all software and systems up to date with the latest security patches and updates. Attackers often exploit known vulnerabilities in outdated software.

Conduct regular vulnerability assessments: Conduct regular vulnerability assessments to identify any weaknesses in the network or systems that could be exploited by attackers.

Train employees: Train employees to identify and report any suspicious activity they observe, such as phishing emails or unauthorized access attempts.

Monitor social media and other sources: Monitor social media and other online sources for any information that may indicate a potential attack, such as a hacker group claiming responsibility for a breach or an announcement of a new exploit.

By implementing these practices, you can help identify suspicious activity before it becomes a complete attack and take proactive measures to protect your network and systems.

**2. *Can you provide the best ML technique used for this purpose and its implementation?***

There is no one-size-fits-all approach when it comes to detecting suspicious network activity using machine learning, as the best technique will depend on the specific problem and data at hand. However, here are some common machine learning techniques that have been used for this purpose:

Supervised learning: Supervised learning algorithms can be used to train a model on labeled data to classify network traffic as either normal or suspicious. Common supervised learning algorithms used for this purpose include decision trees, random forests, support vector machines (SVMs), and deep learning algorithms such as convolutional neural networks (CNNs).

Unsupervised learning: Unsupervised learning algorithms can be used to detect anomalies in network traffic by identifying patterns that deviate from normal behavior. Common unsupervised learning algorithms used for this purpose include clustering algorithms such as k-means, hierarchical clustering, and density-based clustering.

Reinforcement learning: Reinforcement learning algorithms can be used to learn optimal security policies for protecting a network against attacks. This approach involves training a model to maximize a reward signal, such as minimizing false positives or false negatives.

Implementation of these techniques will involve several steps, including data preprocessing, feature engineering, model selection, training, and evaluation. Here's a high-level overview of the implementation process:

Data preprocessing: Prepare the dataset by cleaning, normalizing, and transforming the data to ensure it is suitable for machine learning algorithms.

**INSE 6630 – Recent Developments in Information Systems Security**  
**Assignment 2. Due: 11:59 pm, Sunday, April 16, 2023.**

---

Feature engineering: Extract relevant features from the dataset that can be used to train the model. These may include features such as source and destination IP addresses, port numbers, packet size, payload, and time of day.

Model selection: Choose an appropriate machine learning algorithm for the specific problem at hand. This may involve evaluating different algorithms using performance metrics such as accuracy, precision, recall, and F1 score.

Training: Train the selected model on the labeled data using an appropriate training algorithm, such as stochastic gradient descent or backpropagation.

Evaluation: Evaluate the performance of the trained model using appropriate metrics such as accuracy, precision, recall, and F1 score. You may also use techniques such as cross-validation to ensure the model is robust to different datasets.

Deployment: Once you have a trained and evaluated model, deploy it to the target environment and integrate it into your security infrastructure to detect suspicious network activity in real-time.

Overall, the implementation of machine learning techniques for detecting suspicious network activity requires a combination of domain expertise, data science skills, and software engineering skills. It's important to carefully evaluate the performance of the models and continuously update them as new data becomes available.

**f) Different prompts for the same use cases:**

*1. What are the various ML techniques for detecting the suspicious activity in the network?*

There are several machine learning (ML) techniques that can be used for detecting suspicious activity in a network. Here are some of the common ones:

Anomaly detection: Anomaly detection techniques are used to identify network traffic that deviates from normal behavior. This can be done using unsupervised learning algorithms, such as clustering or density-based methods, or supervised learning algorithms, such as support vector machines (SVMs) or decision trees. Anomaly detection can be useful for identifying new or unknown attacks.

Classification: Classification techniques are used to classify network traffic as either normal or suspicious. This can be done using supervised learning algorithms, such as logistic regression, SVMs, or neural networks. Classification can be useful for identifying known types of attacks.

Clustering: Clustering techniques are used to group network traffic into similar patterns or clusters. This can be done using unsupervised learning algorithms, such as k-means or hierarchical clustering. Clustering can be useful for identifying similar types of attacks or anomalies.

Deep learning: Deep learning techniques, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), can be used for detecting suspicious

**INSE 6630 – Recent Developments in Information Systems Security**  
**Assignment 2. Due: 11:59 pm, Sunday, April 16, 2023.**

---

activity in network traffic. Deep learning can be useful for detecting complex patterns in network traffic and for classifying different types of attacks.

**Ensemble methods:** Ensemble methods combine multiple ML models to improve performance and reduce the risk of overfitting. This can be done using techniques such as bagging, boosting, or stacking.

**Reinforcement learning:** Reinforcement learning can be used to train agents to learn how to respond to different types of network attacks. This can be useful for developing adaptive and dynamic defense strategies.

It's important to note that the choice of ML technique will depend on the specific problem and data at hand, as well as the resources and expertise available for implementation. A combination of different techniques may also be used to improve the overall performance of the system.

**2. *What are the best features for identifying suspicious activity in the network?***

The best features for identifying suspicious activity in a network will depend on the specific problem and data at hand. However, here are some common features that are often used in network traffic analysis:

**Source and destination IP addresses:** The IP addresses of the source and destination of network traffic can provide information about the origin and destination of the traffic.

**Protocol type:** The type of protocol used in network traffic, such as TCP or UDP, can provide information about the nature of the traffic.

**Port numbers:** The port numbers used in network traffic can provide information about the type of service or application being used.

**Packet size:** The size of the packets used in network traffic can provide information about the type of traffic and the nature of the communication.

**Payload:** The payload of the packets used in network traffic can provide information about the content of the communication.

**Time of day:** The time of day when network traffic occurs can provide information about normal behavior patterns and identify anomalies.

**Network flow:** The characteristics of network flow, such as the duration and volume of traffic, can provide information about the nature of the communication.

**Domain names:** The domain names used in network traffic can provide information about the services or applications being accessed.

**DNS queries:** DNS queries can provide information about the domain names being accessed and can help identify suspicious activity.

**User-agent:** The user-agent used in network traffic can provide information about the software or device used to access the network.

These features can be used in combination with machine learning algorithms to detect suspicious activity in network traffic. It's important to note that the choice of features will depend on the specific problem at hand and may require domain expertise in network security. Additionally, feature selection and engineering techniques may also be used to extract the most informative features for the problem at hand.

**INSE 6630 – Recent Developments in Information Systems Security**  
**Assignment 2. Due: 11:59 pm, Sunday, April 16, 2023.**

---

**Differences:**

Prompt Number	First set of Prompts	Second set of Prompts
1	It specifies the various steps involved in identifying the suspicious activity.	It explains about the various ML techniques used in the detection process.
2	It describes the best ML technique for the suspicious activity detection and its implementation.	It specifies the best features which can be used in the implementation of the machine learning.

**Comment:**

- The first series of prompts is more general in nature and does not provide the precise information that we require.
- The second group of responses produced the precise responses needed to identify the problems.

In my opinion, the **second set of the queries** provides the implementation and helps us in identifying the necessary features.