CS973: Machine Learning for Cyber Security

Quiz II (27 Oct 2024)

Q3. Cuckoo in a Coalmine **(3 marks)**

For what type of analysis is the Cuckoo sandbox used? Tick all options that apply (multiple options may be correct).

[ ] Static Analysis

[✓] Dynamic Analysis

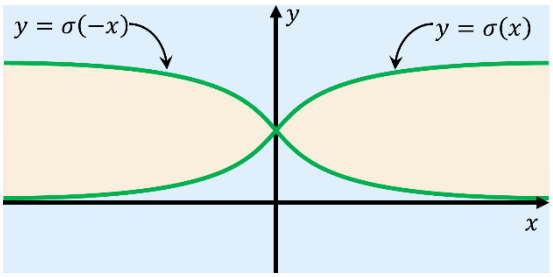
[✓] Hybrid Analysis

[ ] None of the above

Justify your answer briefly below in one-or-two sentences describing what utility does a VM sandbox environment like Cuckoo provide with respect to malware analysis.

VM sandboxes such as Cuckoo allow malware candidates to be executed in a safe and confined environment and investigate their behavioural aspects such as memory actions taken by the executable, file accesses, file movements, network connections established, insertion into autorun or startup lists, attempting logs e.g. keyboard log or network log. These analyses cannot be carried out easily by inspecting the binary or disassembling the binary.

Q4. A Sigmoidal Cross **(4 marks)**



The figure above depicts a binary classification task we wish to solve. The green lines denote the curves $y = \sigma(x)$ and $y = \sigma(-x)$ where $\sigma(x) = (1 + \exp(-x))^{-1}$ is the sigmoid function. Create a feature map $\phi: \mathbb{R}^2 \to \mathbb{R}^D$ for some integer $D > 0$ so that for any 2D vector $\mathbf{z} = (x, y) \in \mathbb{R}^2$, the value of $\text{sign}\left(\mathbf{1}^\top \phi(\mathbf{z})\right)$ is $+1$ if $\mathbf{z}$ is in the yellow region and -1 if $\mathbf{z}$ is in the blue region. The D-dimensional all-ones vector is denoted as $\mathbf{1} = (1,1,\ldots,1) \in \mathbb{R}^D$. Write

down your feature map in the space below. In order to create your feature map, you are allowed to use common functions such as polynomials, absolute value function, exponential function and even sigmoid function i.e., a feature map of the following kind would be valid (although it may not solve the problem) $\phi(\mathbf{z}) = \big(x, y, \exp(x-y), y^2 - x^2, \sigma(y)\big)$.

Notice that the blue region covers exactly those regions that are either above both curves or below both curves i.e. either when $y \geq \max\{\sigma(x), \sigma(-x)\}$ or else $y \leq \min\{\sigma(x), \sigma(-x)\}$. For such points, $\big(y - \sigma(x)\big)\big(y - \sigma(-x)\big) \geq 0$ i.e., the decision boundary is $y^2 - y\big(\sigma(x) + \sigma(-x)\big) + \sigma(x)\sigma(-x) = 0$. However, since we want the $+1$ label in the yellow region and not the blue region, we need to change the orientation of the decision boundary by negating the expression to get $-y^2 + y\big(\sigma(x) + \sigma(-x)\big) - \sigma(x) \cdot \sigma(-x) = 0$. Thus, the following 4D feature map would work

$$\phi(\mathbf{z}) = \big(-y^2, y\sigma(x), y\sigma(-x), -\sigma(x) \cdot \sigma(-x)\big)$$

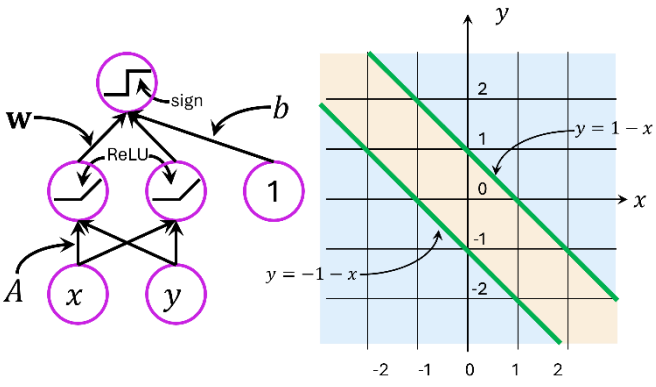
One can simplify the map by noticing that $\sigma(-x) = 1 - \sigma(x)$ to get

$$\phi(\mathbf{z}) = \big(-y^2, y, -\sigma(x), \big(\sigma(x)\big)^2\big)$$

Variants such as the 3D map $\phi(\mathbf{z}) = \big(-y^2, y\big(\sigma(x) + \sigma(-x)\big), -\sigma(x) \cdot \sigma(-x)\big)$ or even the 1D map $\phi(\mathbf{z}) = \big(-y^2 + y - \sigma(x) + \big(\sigma(x)\big)^2\big)$ are correct too.

Q5. NN hack **(7 marks)**

We wish to use a neural network, with network diagram given above to the left, to solve a binary classification problem depicted on the right. The green decision boundaries in the figure are the lines $x + y = 1$ and $x + y = -1$. The neural network has as parameters a 2×2 matrix $A \in \mathbb{R}^{2\times 2}$, a 2D vector $\mathbf{w} \in \mathbb{R}^2$ and a bias term $b \in \mathbb{R}$. The output of the neural network is $\text{sign}(\mathbf{w}^\top \phi(\mathbf{x}) + b)$ where $\phi(\mathbf{x}) = \max(A\mathbf{x}, 0)$ where the max operation is applied coordinate-wise (i.e.



The ReLU activation). Find out values of the parameters A, $\mathbf{w}$, b so that the neural network gives output $+1$ in the yellow region and -1 in the blue region.

The classifier desired is $\text{sign}(1 - |x + y|)$ (and not $\text{sign}(|x + y| - 1)$) as we want the $+1$ label in the yellow region and -1 in the blue region. As we have seen in the lecture videos, $|t| = \text{ReLU}(t) + \text{ReLU}(-t)$ i.e. the classifier is

$$\text{sign}\big(1 - \text{ReLU}(x + y) + \text{ReLU}(-x - y)\big)$$

This tells us that $\mathbf{w} = (-1, -1), b = 1$ and $A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$

Q6. The Melbo Equation **(6 marks)**

If Melbo has studied diligently for an exam, then with probability 75%, Melbo gets full marks in that exam i.e., with 25% probability, Melbo does not score full marks despite having studied diligently. If Melbo has not studied diligently for an exam, then Melbo can never get full marks in that exam. Now, Melbo is a very busy person with other things to do in life besides studying for exams (such as being the star of YouTube videos). Thus, for any exam, Melbo decides to study diligently for that exam with probability 50% (this choice is independent of what Melbo did for previous exams). It is known that Melbo scored full marks in the CS973 exam but did not score full marks in the CS771 exam.

Let S denote the event that Melbo studied diligently and let F denote the event that Melbo gets full marks. We are told that $\mathbb{P}[F \mid S] = 0.75$ and $\mathbb{P}[F \mid \neg S] = 0$ as well as that $\mathbb{P}[S] = 0.5$. By law of total probability, this tells us that

$$\mathbb{P}[F] = \mathbb{P}[F \mid S] \cdot \mathbb{P}[S] + \mathbb{P}[F \mid \neg S] \cdot \mathbb{P}[\neg S] = 0.75 \cdot 0.5 = 0.375 = \frac{3}{8}$$

Answer the 3 parts of the question below in the space provided. Given only the final answers for each part either as a fraction or a decimal or a percentage -- no derivations required.

What is the probability that Melbo studied diligently for the CS973 exam?

Since Melbo could not have obtained full marks without studying diligently, we must have $\mathbb{P}[S \mid F] = 1$. Another way of obtaining the same answer is to use the Bayes rule to get

$$\mathbb{P}[S \mid F] = \frac{\mathbb{P}[F \mid S] \cdot \mathbb{P}[S]}{\mathbb{P}[F]} = \frac{0.75 \cdot 0.5}{0.375} = 1$$

What is the probability that Melbo studied diligently for the CS771 exam?

Since $\mathbb{P}[F \mid S] = 0.75$, we have $\mathbb{P}[\neg F \mid S] = 0.25$. Applying the Bayes rule gives

$$\mathbb{P}[S \mid \neg F] = \frac{\mathbb{P}[\neg F \mid S] \cdot \mathbb{P}[S]}{\mathbb{P}[\neg F]} = \frac{0.25 \cdot 0.5}{1 - 0.375} = 0.2 = \frac{1}{5}$$

Melbo has a third exam for CS315 coming up next week. What is the probability that Melbo will score full marks in that exam?

As calculated before, we have

$$\mathbb{P}[F] = 0.375 = \frac{3}{8}$$