Answer to Exercise 1

Since $f(1,3) = \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 0.5 = -1 + 0.5 = -0.5 < 0$, the point $(1,3)$ is initially classified into Class 2, and the corresponding adversarial sample $x' = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix}$ should be classified into Class 1, i.e., $f(x_1', x_2') > 0$.

$$\frac{\partial f}{\partial x} = \frac{\partial W \cdot x + b}{\partial x} = W = \begin{bmatrix} 2 & -1 \end{bmatrix}$$

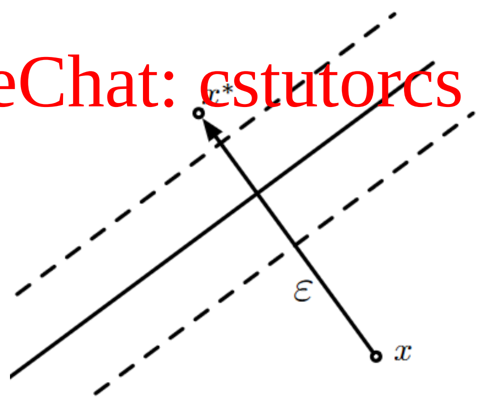
Step 1:

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix} \leftarrow Clip_3\left(\begin{bmatrix} 1 \\ 3 \end{bmatrix} + 1 \cdot sgn\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right)\right)$$

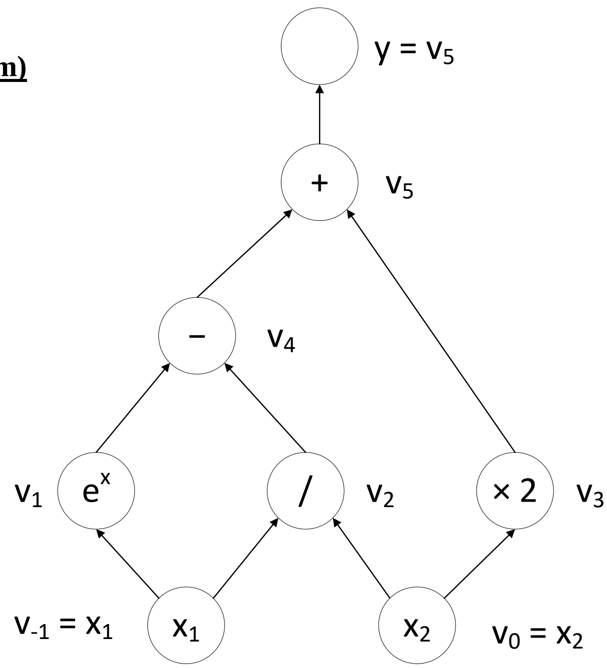
Since $f(2,2) = \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} + 0.5 = 2 + 0.5 = 2.5 > 0$, i.e., $(2,2)$ is classified into the other class, it is the adversarial sample for $(1,3)$.

Note:

In this case, in order to generate an adversarial sample, we perturb a given input in the direction orthogonal to the decision boundary hyperplane.



Answer to Exercise 2



Forward evaluation trace **<u>(start from top to bottom)</u>**

$$
\begin{aligned}
&v_{-1} = x_1 = 2\\
&v_0 = x_2 = 4\\
&v_1 = e^{v_{-1}} = e^2\\
&v_2 = v_{-1}/v_0 = 0.5\\
&v_3 = 2 \cdot v_0 = 8\\
&v_4 = v_1 - v_2 = e^2 - 0.5\\
&v_5 = v_3 + v_4 = e^2 + 7.5\\
&y = v_5
\end{aligned}
$$

Forward derivative trace

1. For calculating $\frac{\partial y}{\partial x_1}$ **<u>(start from top to bottom)</u>**

$$
\begin{aligned}
&\dot{v}_{-1} = \dot{x}_1 = 1\\
&\dot{v}_0 = \dot{x}_2 = 0\\
&\dot{v}_1 = e^{v_{-1}} \cdot \dot{v}_{-1} = e^2\\
&\dot{v}_2 = \frac{\dot{v}_{-1} \cdot v_0 - v_{-1} \cdot \dot{v}_0}{{v_0}^2} = \frac{1 \times 4 - 2 \times 0}{16} = \frac{1}{4}\\
&\dot{v}_3 = 2 \cdot \dot{v}_0 = 0\\
&\dot{v}_4 = \dot{v}_1 - \dot{v}_2 = e^2 - \frac{1}{4}\\
&\dot{v}_5 = \dot{v}_3 + \dot{v}_4 = e^2 - \frac{1}{4}\\
&\dot{y} = \dot{v}_5 = e^2 - \frac{1}{4}
\end{aligned}
$$

2. For calculating $\frac{\partial y}{\partial x_2}$ **<u>(start from top to bottom)</u>**

$$
\begin{aligned}
&\dot{v}_{-1} = \dot{x}_1 = 0\\
&\dot{v}_0 = \dot{x}_2 = 1\\
&\dot{v}_1 = e^{v_{-1}} \cdot \dot{v}_{-1} = 0\\
&\dot{v}_2 = \frac{\dot{v}_{-1} \cdot v_0 - v_{-1} \cdot \dot{v}_0}{{v_0}^2} = \frac{0 \times 4 - 2 \times 1}{16} = -\frac{1}{8}\\
&\dot{v}_3 = 2 \cdot \dot{v}_0 = 2\\
&\dot{v}_4 = \dot{v}_1 - \dot{v}_2 = \frac{1}{8}\\
&\dot{v}_5 = \dot{v}_3 + \dot{v}_4 = 2\frac{1}{8}\\
&\dot{y} = \dot{v}_5 = 2\frac{1}{8}
\end{aligned}
$$

Reverse adjoint trace <u>(start from bottom to top)</u>

$\bar{x}_1 = \bar{v}_{-1} = e^2 - \frac{1}{4}$

$\bar{x}_2 = \bar{v}_0 = 2\frac{1}{8}$

$\bar{v}_{-1} = \bar{v}_{-1} + \bar{v}_1 \cdot \frac{\partial v_1}{\partial v_{-1}} = -\frac{1}{4} + 1 \times e^{v_{-1}} = -\frac{1}{4} + e^2$

$\bar{v}_0 = \bar{v}_0 + \bar{v}_2 \cdot \frac{\partial v_2}{\partial v_0} = 2 + (-1) \times \left(-\frac{v_{-1}}{{v_0}^2}\right) = 2 + \frac{1}{8}$

$\bar{v}_{-1} = \bar{v}_2 \cdot \frac{\partial v_2}{\partial v_{-1}} = -1 \times \frac{1}{v_0} = -\frac{1}{4}$

$\bar{v}_0 = \bar{v}_3 \cdot \frac{\partial v_3}{\partial v_0} = 2$

$\bar{v}_2 = \bar{v}_4 \cdot \frac{\partial v_4}{\partial v_2} = -1$

$\bar{v}_1 = \bar{v}_4 \cdot \frac{\partial v_4}{\partial v_1} = 1$

$\bar{v}_3 = \bar{v}_5 \cdot \frac{\partial v_5}{\partial v_3} = 1$

$\bar{v}_4 = \bar{v}_5 \cdot \frac{\partial v_5}{\partial v_4} = 1$

$\bar{v}_5 = \bar{y} = 1$