COMP90073 – Security Analytics

Week 12 Tutorial

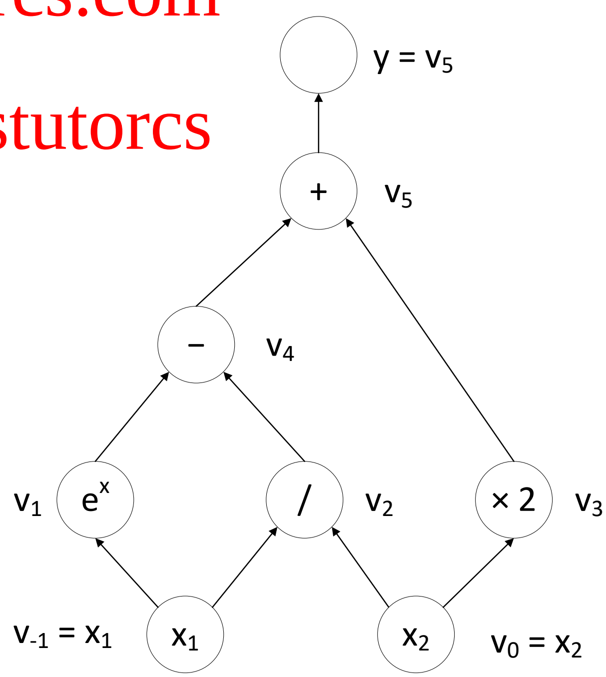
1. Suppose that f is a binary linear classifier $f(x; W, b) = W \cdot x + b$, where $W = [2 \quad -1], b = 0.5$, and $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, i.e., the input x is two dimensional. Given a point $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, it will be classified into Class 1 if $f(x) > 0$, or Class 2 otherwise. For example,

(1) since $f(2, 1) = [2 \quad -1] \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 0.5 = 3 + 0.5 = 3.5 > 0$, the point $(2, 1)$ is classified into Class 1;

(2) since $f(-1, 1) = [2 \quad -1] \begin{bmatrix} -1 \\ 1 \end{bmatrix} + 0.5 = -3 + 0.5 = -2.5 < 0$, the point $(-1, 1)$ is classified into Class 2;

Generate the adversarial sample for point $(1, 3)$ using the iterative gradient sign method. The parameters in this algorithm are given as follows: (1) the step size is fixed to 1, (2) $\epsilon = 3$ – the intermediate and final results need to be clipped if necessary, to make sure that they are in the ϵ-neighbourhood of the original point, i.e., $|x_i - x_i'| \leq \epsilon, i = 1, 2$.

2. Use automatic differentiation to calculate the partial derivatives $\left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}\right)$ for $y = e^{x_1} - x_1/x_2 + 2x_2$ at point (2, 4).



Forward evaluation trace

$v_{-1} = x_1$
$v_0 = x_2$
$v_1 =$ ________
$v_2 =$ ________
$v_3 =$ ________
$v_4 =$ ________
$v_5 =$ ________
$y = v_5$

Forward derivative trace

(1) For calculating $\frac{\partial y}{\partial x_1}$

$\dot{v}_{-1} = \dot{x}_1$
$\dot{v}_0 = \dot{x}_2$
$\dot{v}_1 =$ ________
$\dot{v}_2 =$ ________
$\dot{v}_3 =$ ________
$\dot{v}_4 =$ ________

$\dot{v}_5 =$ ________
$\dot{y} = \dot{v}_5$

(2) For calculating $\frac{\partial y}{\partial x_2}$

$\dot{v}_{-1} = \dot{x}_1$
$\dot{v}_0 = \dot{x}_2$
$\dot{v}_1 =$ ________
$\dot{v}_2 =$ ________
$\dot{v}_3 =$ ________
$\dot{v}_4 =$ ________
$\dot{v}_5 =$ ________
$\dot{y} = \dot{v}_5$

Reverse adjoint trace

$\bar{x}_1 = \bar{v}_{-1}$
$\bar{x}_2 = \bar{v}_0$
$\bar{v}_{-1} =$ ________
$\bar{v}_0 =$ ________
$\bar{v}_{-1} =$ ________
$\bar{v}_0 =$ ________
$\bar{v}_2 =$ ________
$\bar{v}_1 =$ ________
$\bar{v}_3 =$ ________
$\bar{v}_4 =$ ________
$\bar{v}_5 = \bar{y}$