Problem on LIME.

* black box, Complex Model : Logistic Regression ($f(x)$)
  Explainable Model : linear Regression ($g(x)$)

* features : x_1 (Age) x_2 (Income)

* weights : $w_1 \rightarrow 0.5$, $w_2 \rightarrow 0.3$; bias $b \rightarrow -2$ (Given)
  $\theta_1 = 0.2$ $\theta_2 = 0.5$ $\theta_0 = \cancel{0.7}\ 0.3$ (Given)

Logistic Regression (Complex Model)

$$\boxed{f(x) = \frac{1}{1+e^{-(w_1x_1 + w_2x_2 + b)}}}$$

linear Model (Surrogate) → linear Regression

$$\boxed{g(x) = \theta_1 x_2 + \theta_2 x_1 + \theta_0}$$

① $\boxed{x_1 = 30,\ x_2 = 50.}$

② Generate perturbed data points.

	x_1	x_2
	30.5	50.2
x_2	29.7	50.5
x_3	30.2	49.8
x_4	30.1	50.3
x_5	15	18.2

③ Compute the prediction for perturbed datapoints.

$$\boxed{f(x) = \frac{1}{1+e^{-(w_1x_1 + w_2x_2 + b)}}}$$

$$= \frac{1}{1+e^{-(0.5 \times 30.5 + 0.3 \times 50.2 - 2)}}$$

$$= \frac{1}{1+e^{-(15.25 + 15.06 - 2)}} = \frac{1}{1+e^{-28.31}}$$

$$= 0.9999$$

2) $$f(x) = \frac{1}{1+e^{-(0.5\times 29.7 + 0.3\times 50.5 - 2)}}$$

$$= \frac{1}{1+e^{-(14.85+15.15-2)}} = \frac{1}{1+e^{-28}}$$

$$= \boxed{0.9999}$$

3) $$f(x) = \frac{1}{1+e^{-(0.5\times 30.2 + 0.3\times 49.8 - 2)}}$$

$$= \frac{1}{1+e^{-(15.2+14.94-2)}} = \frac{1}{1+e^{-28.44}} = \boxed{0.9999}$$

4) $$f(x) = \frac{1}{1+e^{-(0.5\times 30.1 + 0.3\times 50.3 - 2)}} = \frac{1}{1+e^{-28.14}}$$

$$= \frac{1}{1+e^{-(15.05+15.09-2)}} = \boxed{0.9999}$$

5) $$f(x) = \frac{1}{1+e^{-(0.5\times 15 + 0.3\times 18.2 - 2)}} = \frac{1}{1+e^{-10.96}}$$

$$\frac{1}{1+e^{-(7.5+5.46-2)}} = \boxed{0.9999}$$

(4) Calculate $G(x)$ Linear Regression.

$$\boxed{g(x) = \theta_1 x_1 + \theta_2 x_2 + \theta_0}$$

Given $\theta_1 = 0.2$ $\quad \theta_2 = 0.5$ $\quad \theta_0 = 0.3$

1) $0.2\times 30.5 + 0.5\times 50.2 + 0.3 = 6.1 + 25.1 + 0.3 = 31.5$

2) $0.2\times 29.7 + 0.5\times 50.5 + 0.3 = 5.94 + 25.25 + 0.3 = 31.49$

3) $0.2\times 30.2 + 0.5\times 49.8 + 0.3 = 6.04 + 24.9 + 0.3 = 31.24$

4) $0.2\times 30.1 + 0.5\times 50.3 + 0.3 = 6.02 + 25.15 + 0.3 = 31.47$

5) $0.2\times 15 + 0.5\times 18.3 + 0.3 = 3 + 9.15 + 0.3 = 12.45$

$\boxed{\sigma^2 = 1}$ In std Normal distribution.

Mean = 0

σ^2(Var) = 1

⑤ Assign weights to perturbed data points.

$$w_i = \exp -\left(\frac{\Sigma_{ij} |x_1 - x_2|^2}{2\sigma^2} \right)$$

① $$= e^{-\left(\frac{|30.5-30|^2 + |50.2-50|^2}{2} \right)}$$

$$= e^{-\left(\frac{0.5^2 + 0.2^2}{2} \right)} = e^{-\frac{0.25+0.04}{2}} = e^{-\frac{0.29}{2}}$$

$$= e^{-0.145} = \boxed{0.865}$$

② $$e^{-\left(\frac{|29.7-30|^2 + |50.5-50|^2}{2} \right)} = e^{-\left(\frac{0.09+0.25}{2} \right)}$$

$$= e^{-\left(\frac{0.34}{2} \right)} = e^{-0.17} = \boxed{0.843}$$

③ $$e^{-\left(\frac{|30.2-30|^2 + |49.8-50|^2}{2} \right)} = e^{-\left(\frac{0.04+0.04}{2} \right)}$$

$$= e^{-(0.04)} = \boxed{0.960}$$

④ $$e^{-\left(\frac{|30.1-30|^2 + |50.3-50|^2}{2} \right)} = e^{-\left(\frac{0.1+0.09}{2} \right)}$$

$$= e^{-\left(\frac{0.19}{2} \right)} = \boxed{\cancel{0.909}} \quad \boxed{0.909}$$

⑤ $$e^{-\left(\frac{|15-30|^2 + |18.2-50|^2}{2} \right)} = e^{-\left(\frac{225+1011.2}{2} \right)}$$

$$= e^{-\left(\frac{1236}{2} \right)} = e^{-618.1} = \boxed{3.652\, e^{-269}}$$

Calculate.

(6) Locally weighted loss function.

$$L = \Sigma \omega (g(x) - f(x))^2$$

$$= 0.865 (31.5 - 0.99)^2 + 0.843 (31.49 - 0.99)^2$$

$$+ 0.960 (31.24 - 0.99)^2 + 0.909 (31.47 - 0.99)^2$$

$$+ 3.652 e^{-269} (12.45 - 0.999)^2$$

$$= 0.865 (30.51)^2 + 0.843 (30.5)^2 + 0.960 (30.25)^2$$

$$+ 0.909 (30.48)^2 + 3.65 e^{-269} (11.45)^2$$

$$= 805.193 + 784.200 + 878.46 + 844.48$$

$$+ 7.156 e^{-115}$$

$$= \boxed{3213.33}$$

Note : Threshold value will be given.

if L is less than the threshold value.

g(x) is approximating the complex f(x) model

$\boxed{\text{good}}$

Here it is a worst case scenario