Logistic Regression

classification problems are supervised learning problems in which the outputs are discrete values

Regression

$$y \in \mathbb{R}$$

Classification

$$y \in \{0,1\}$$

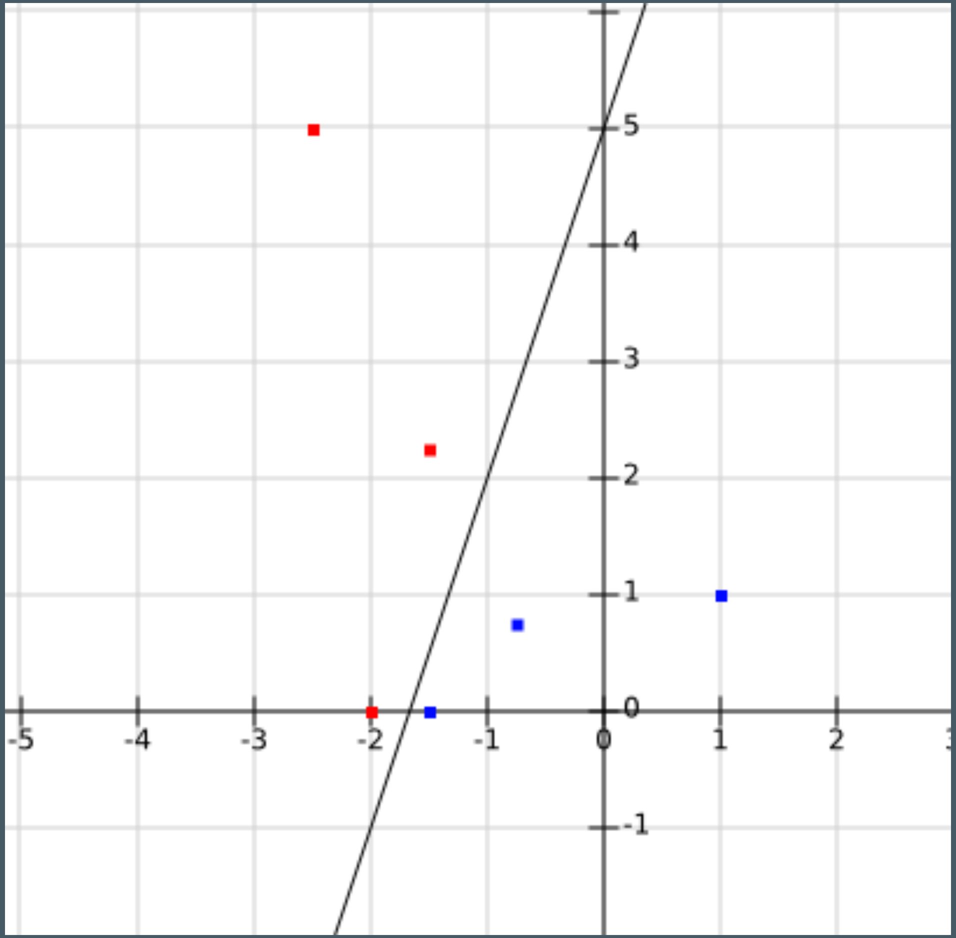
regression problems can be transformed into classification problems by mapping continuous outputs to discrete outputs

$$h_c(x) = \begin{cases} 1 & if \quad h_r(x) \geq 0.5 \\ 0 & if \quad h_r(x) < 0.5 \end{cases}$$

$$h(x) = g(\theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n)$$

$$prediction = \begin{cases} 1 \quad if \quad h(x) \geq 0 \\ 0 \quad if \quad h(x) < 0 \end{cases}$$

Decision Boundary



$$h_{lin}(x) = 5 + 3x_1 - x_2$$

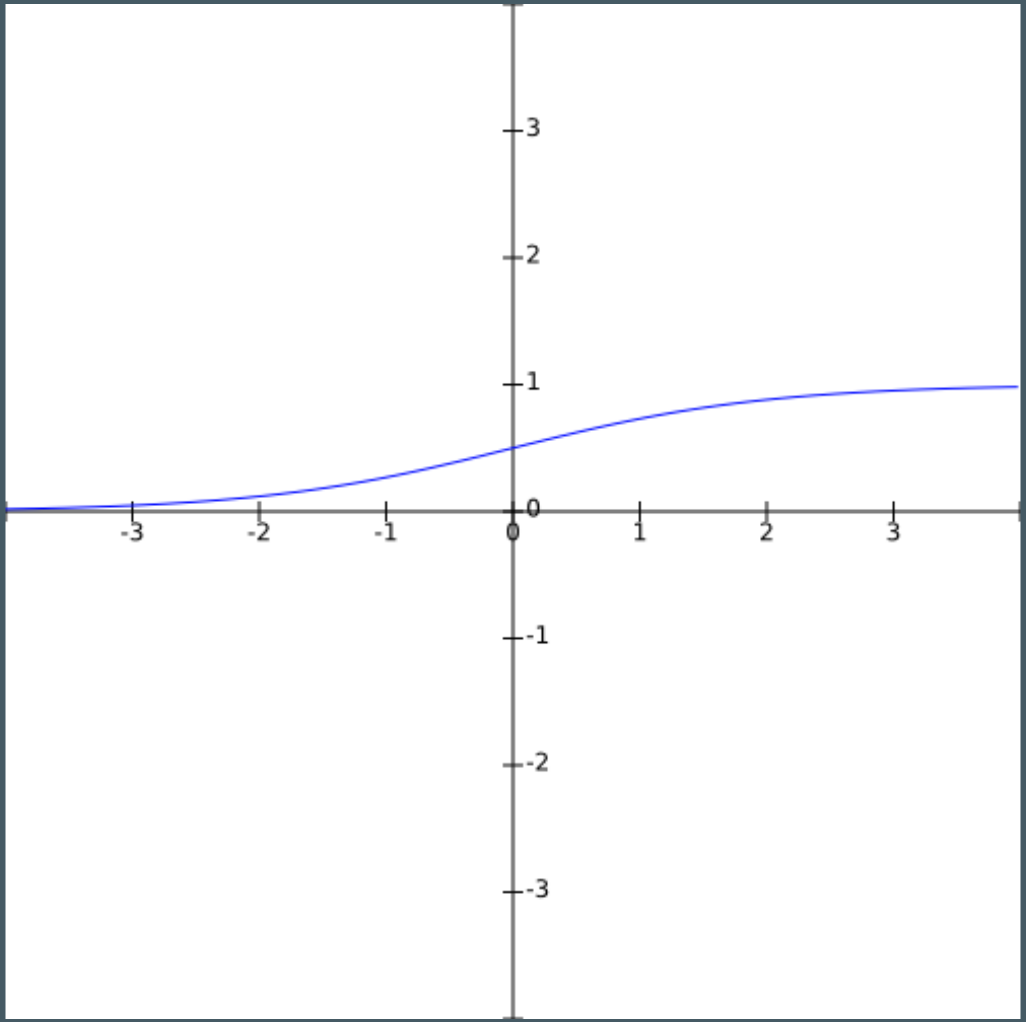
$\boldsymbol{t}$ *example*	$h_{lin}(\boldsymbol{x})$
$(-2.5,5)$	$5+3(-2.5)-5=-7.5$
$(-2,0)$	$5+3(-2)-0=-1$
$(-1.5,2.25)$	$5+3(-1.5)-2.25=-1.75$
$(-0.75,0.75)$	$5+3(-0.75)-0.75=6.5$
$(-1.5,0)$	$5+3(-1.5)-0=0.5$
$(1,1)$	$5+3(1)-1=7$

It doesn't make sense for $h(x)$ to have a range greater than 1 or less than 0 since we know that $y \in \{0,1\}$

So that $h(x)$ satisfies the range

$$0 \leq h(x) \leq 1$$

we make use of an another function to map the values to the correct ranges



Sigmoid Function/Logistic Function

$$g(z) = \frac{1}{1 + e^{-z}}$$

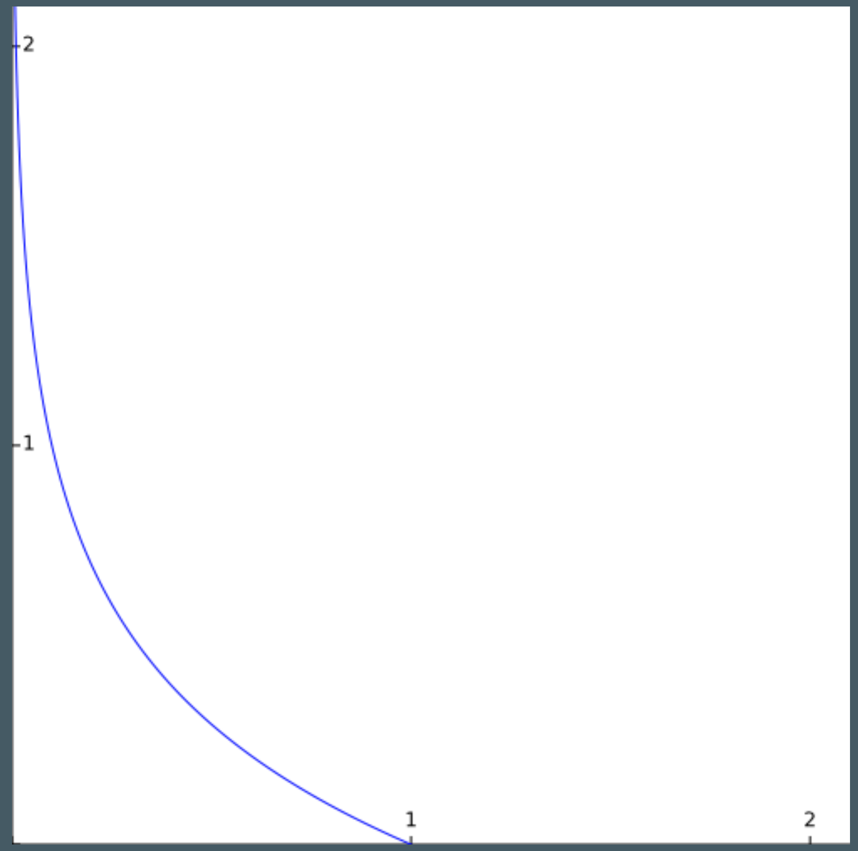
By composing the linear hypothesis into the sigmoid function the logistic hypothesis become more meaningful

The value of the prediction based on the logistic hypothesis can be interpreted as a measure of certainty for the predictor to avoid a non-convex cost function, Logistic regression uses a different cost function compared to the linear regression

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} cost\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right)$$

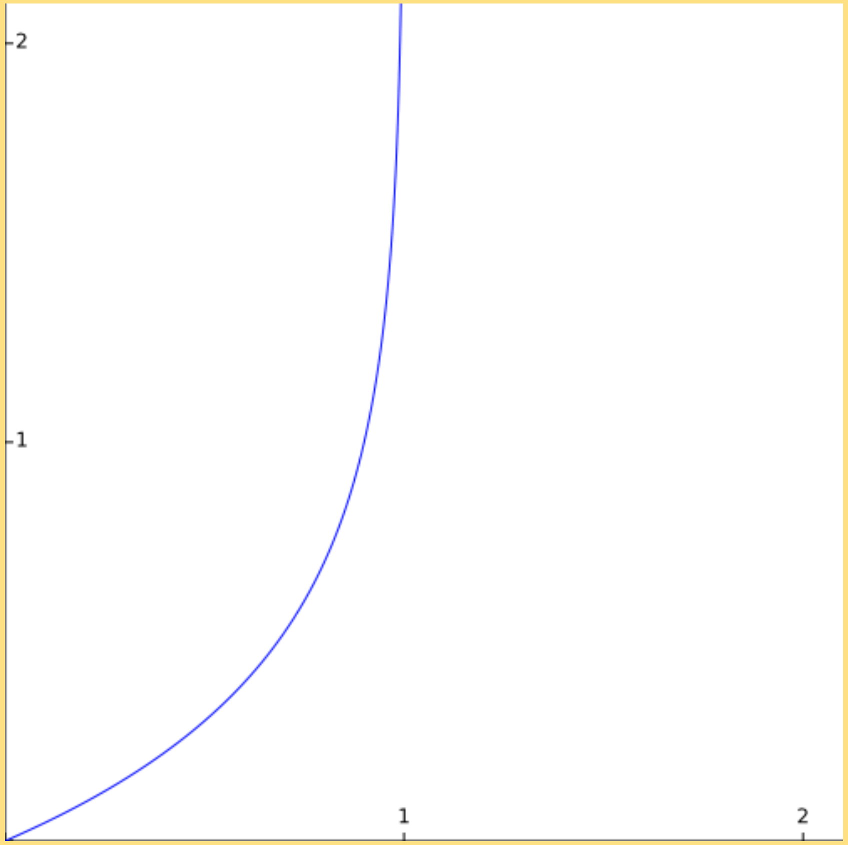
$$cost(h_\theta(x), y) = \begin{cases} -\log\left(h_\theta(x)\right) & if \quad y = 1 \\ -\log\left(1 - h_\theta(x)\right) & if \quad y = 0 \end{cases}$$

$$y = 1$$



$$-\log\big(h_\theta(x)\big)$$

$$y = 0$$



$$-\log\big(1 - h_\theta(x)\big)$$

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta(x^{(i)})) + (1 - y^{(i)})\log(1 - h_\theta(x^{(i)}))]$$

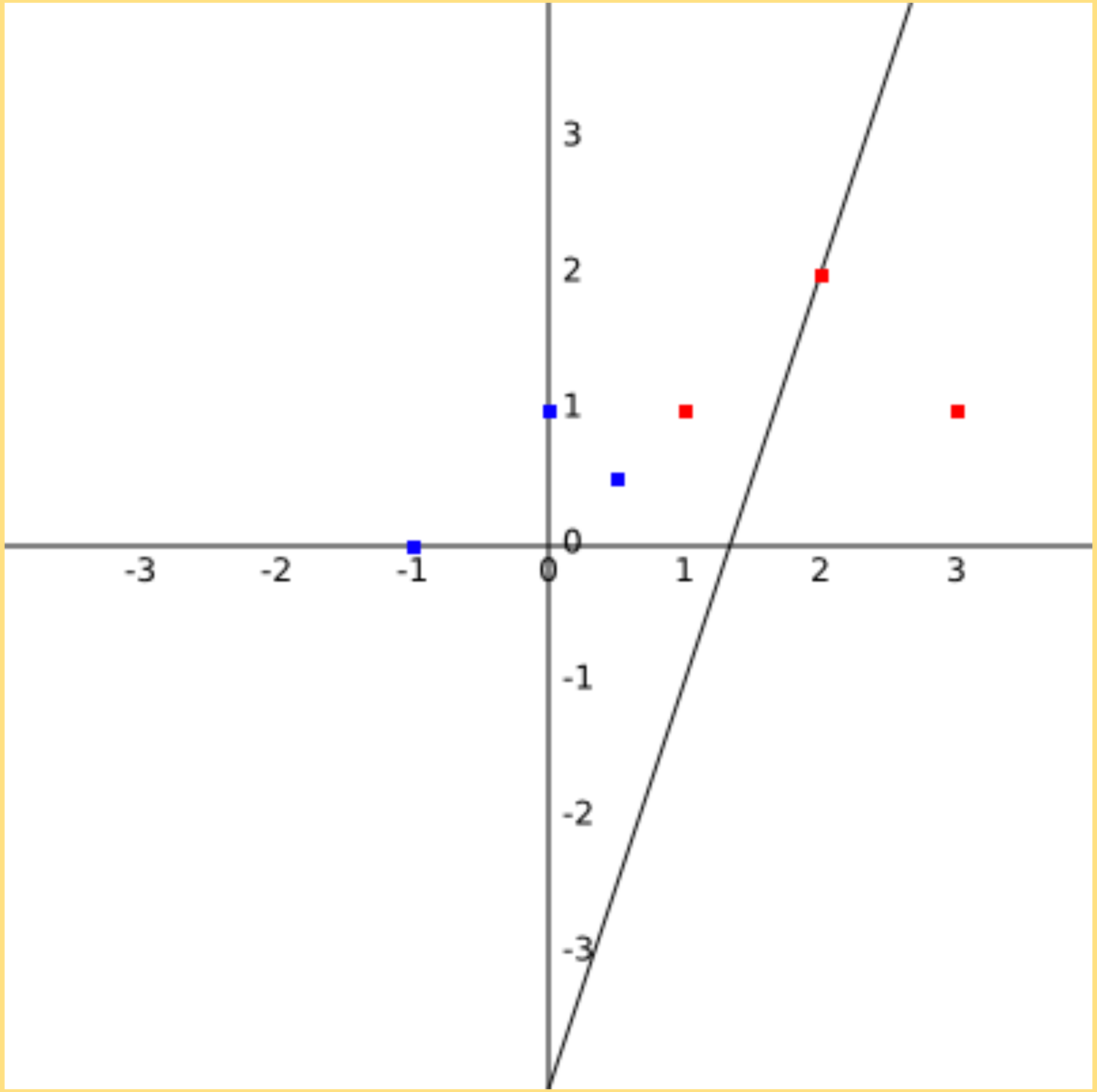
$$J(\theta) = \frac{1}{m}\left[-y^T \log H - (1-y)^T \log(1-H)\right]$$

where: $H = g(X\theta)$

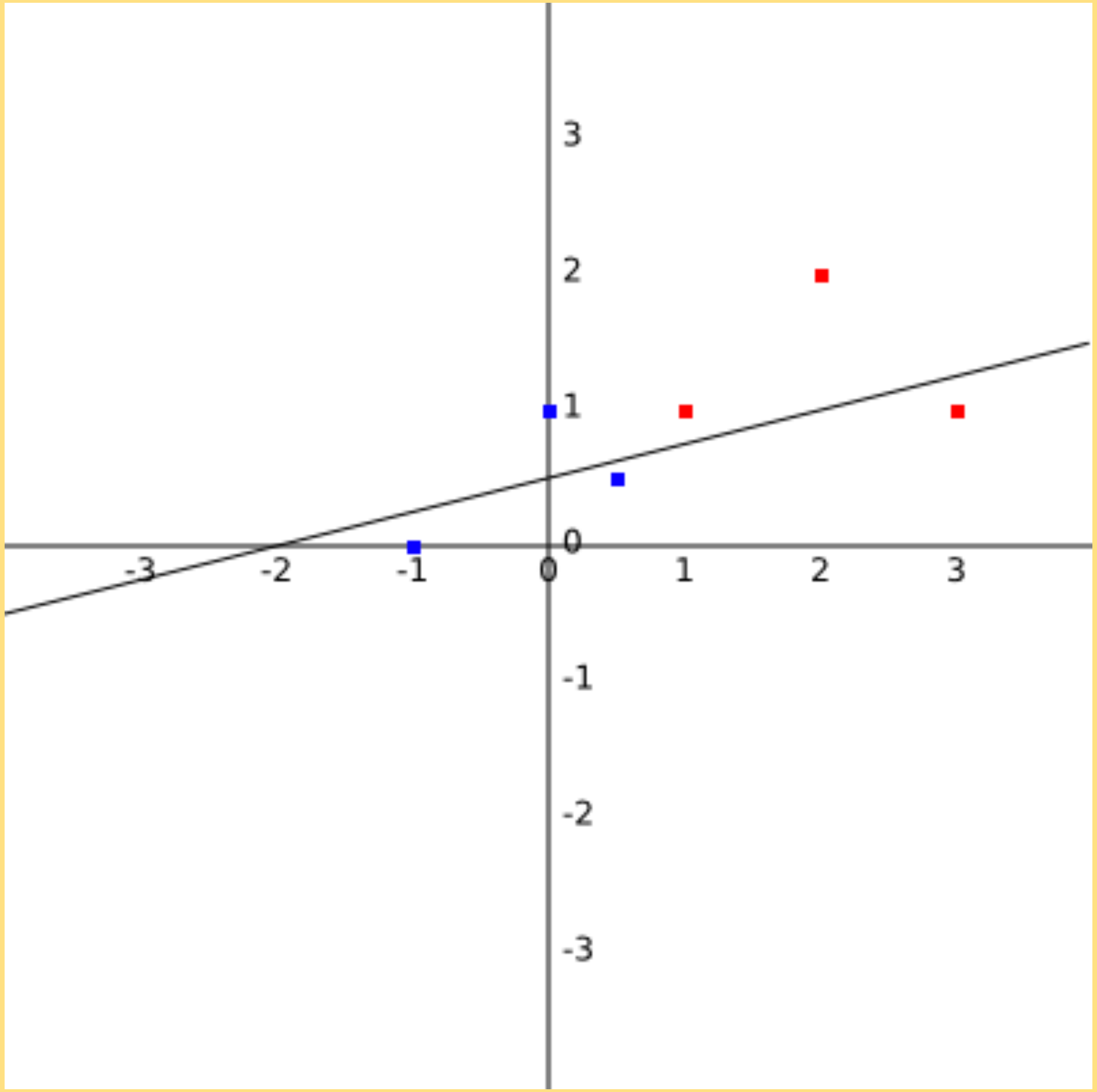
Solve for the cost for the following θ values:

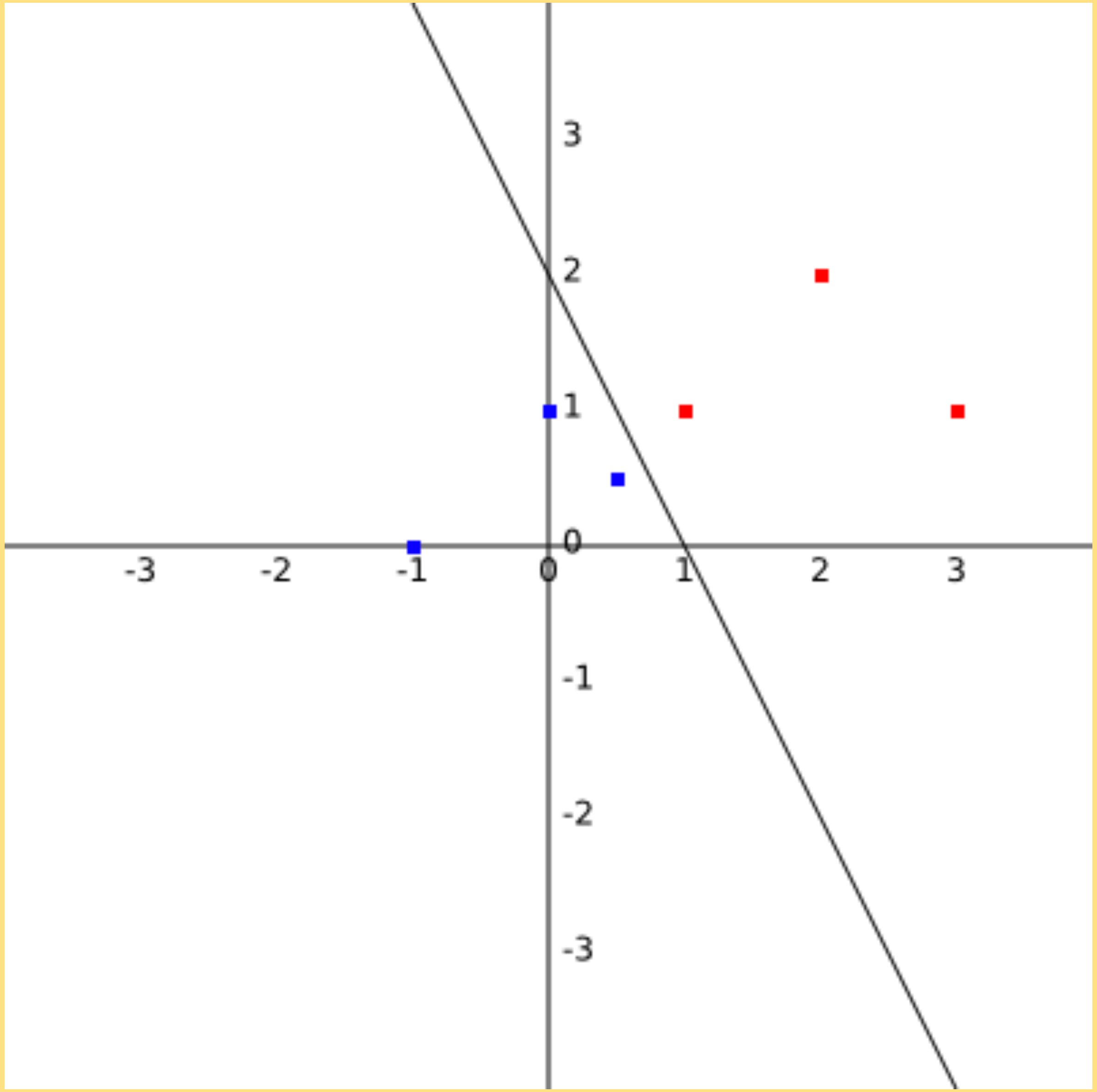
1. $\begin{bmatrix} -4 \\ 3 \\ -1 \end{bmatrix}$ 2. $\begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix}$ 3. $\begin{bmatrix} 8 \\ 4 \\ -8 \end{bmatrix}$

x_1	x_2	y
1	1	1
2	2	1
3	1	1
0.5	0.5	0
0	1	0
-1	0	0



3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3


3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3



3
2
1
0
-1
-2
-3
-3 -2 -1 0 1 2 3

Repeat until convergence$(for\ all\ j\ in\ [0, n])$:{

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)}\right) x_j^{(i)}$$

}

Repeat until convergence:{

$$\theta \ := \theta - \frac{\alpha}{m} X^T (g(\mathrm{X}\theta) - y)$$

}

$$y \in \{\mathrm{Setosa}, \mathrm{Versicolor}\}$$

To apply the logistic regression into real world problems, we assign each **CLASS** into a y value

Sepal Length Cm	Sepal Width Cm	Petal Length Cm	Petal Width Cm	Species (output)	y
5.1	3.5	1.4	0.2	Setosa	1
4.9	3	1.4	0.2	Setosa	1
4.7	3.2	1.3	0.2	Setosa	1
4.6	3.1	1.5	0.2	Setosa	1
7	3.2	4.7	1.4	Versicolor	0
6.4	3.2	4.5	1.5	Versicolor	0
6.9	3.1	4.9	1.5	Versicolor	0
5.5	2.3	4	1.3	Versicolor	0

$$h_{log}(X) = P(y = 1|X) = P(\text{species} = \text{Setosa}|X)$$

Sepal Length Cm	Sepal Width Cm	Petal Length Cm	Petal Width Cm	Species (output)	y
5.1	3.5	1.4	0.2	Setosa	1
4.9	3	1.4	0.2	Setosa	1
4.7	3.2	1.3	0.2	Setosa	1
4.6	3.1	1.5	0.2	Setosa	1
7	3.2	4.7	1.4	Versicolor	0
6.4	3.2	4.5	1.5	Versicolor	0
6.9	3.1	4.9	1.5	Versicolor	0
5.5	2.3	4	1.3	Versicolor	0

$$
\begin{aligned}
&1 - h_{log}(X) \\
&= P(y = 0|X) \\
&= P(\text{species} = \text{Versicolor}|X)
\end{aligned}
$$

$$y \in \{\text{Setosa}, \text{Versicolor}, \text{Virginica}\}$$

Sepal Length Cm	Sepal Width Cm	Petal Length Cm	Petal Width Cm	Species (output)	y
5.1	3.5	1.4	0.2	Setosa	?
4.9	3	1.4	0.2	Setosa	?
4.7	3.2	1.3	0.2	Setosa	?
6.4	3.2	4.5	1.5	Versicolor	?
6.9	3.1	4.9	1.5	Versicolor	?
5.5	2.3	4	1.3	Versicolor	?
6.2	3.4	5.4	2.3	Virginica	?
5.9	3	5.1	1.8	Virginica	?

What numerical value would we use for y?

Sepal Length Cm	Sepal Width Cm	Petal Length Cm	Petal Width Cm	Species (output)	y_1	y_2	y_3
5.1	3.5	1.4	0.2	Setosa	1	0	0
4.9	3	1.4	0.2	Setosa	1	0	0
4.7	3.2	1.3	0.2	Setosa	1	0	0
6.4	3.2	4.5	1.5	Versicolor	0	1	0
6.9	3.1	4.9	1.5	Versicolor	0	1	0
5.5	2.3	4	1.3	Versicolor	0	1	0
6.2	3.4	5.4	2.3	Virginica	0	0	1
5.9	3	5.1	1.8	Virginica	0	0	1

This means that we should train 3 different hypotheses. One for each class

Sepal Length Cm	Sepal Width Cm	Petal Length Cm	Petal Width Cm	Species (output)	y_1	y_2	y_3
5.1	3.5	1.4	0.2	Setosa	1	0	0
4.9	3	1.4	0.2	Setosa	1	0	0
4.7	3.2	1.3	0.2	Setosa	1	0	0
6.4	3.2	4.5	1.5	Versicolor	0	1	0
6.9	3.1	4.9	1.5	Versicolor	0	1	0
5.5	2.3	4	1.3	Versicolor	0	1	0
6.2	3.4	5.4	2.3	Virginica	0	0	1
5.9	3	5.1	1.8	Virginica	0	0	1

$$h_1(X) = P(y_1 = 1|X) \\ = P(\text{species} = \text{Setosa}|X)$$

$$h_2(X) = P(y_2 = 1|X) \\ = P(\text{species} = \text{Versicolor}|X)$$

$$h_3(X) = P(y_3 = 1|X) \\ = P(\text{species} = \text{Virginica}|X)$$

Hypothesis values for multi-class classifications come in the form of matrices. Example:

$$H^{(1)} = \begin{bmatrix} h_1(X^{(1)}) \\ h_2(X^{(1)}) \\ h_3(X^{(1)}) \end{bmatrix}$$

$$\text{Prediction} = \max(h_1(X^{(1)}), h_2(X^{(1)}), h_3(X^{(1)}))$$