

Introduction to Neural Network

Instructors:

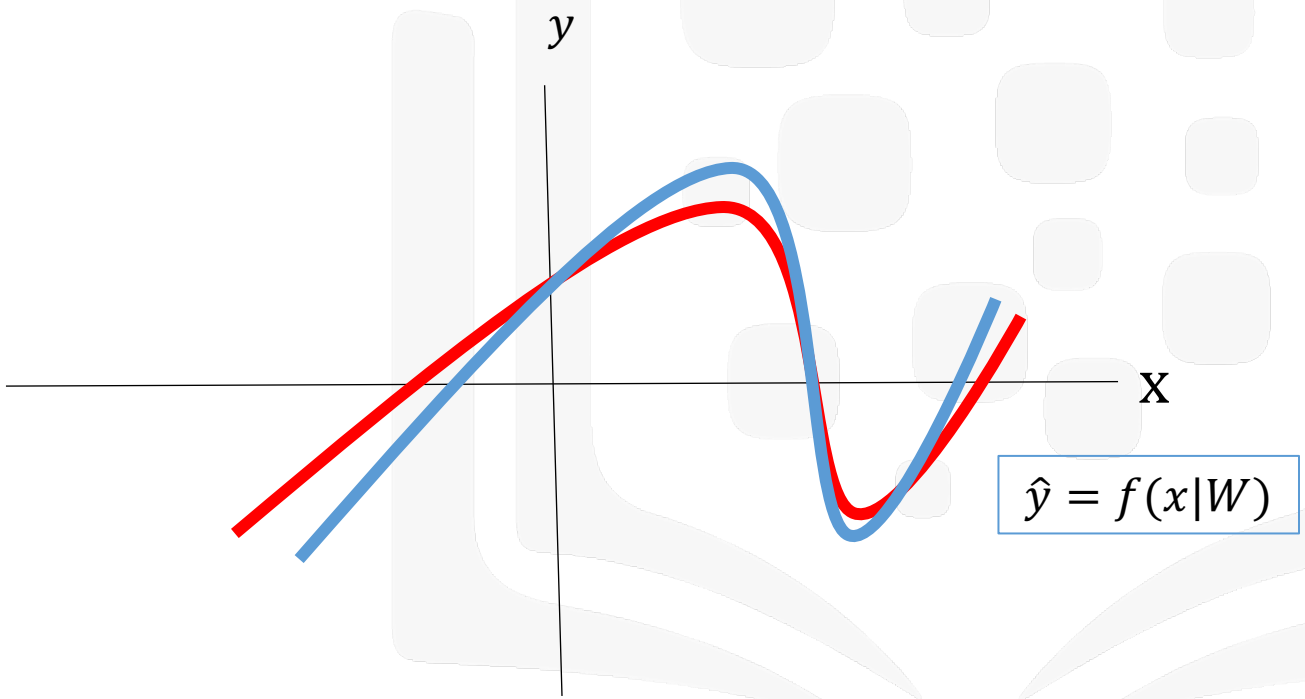
Joseph Santarcangelo, Ph.D.

.

BIG DATA UNIVERSITY

What Is a Neural Network

- A neural network is a function that can use to approximate ‘something’ using a set of parameters



BIG DATA UNIVERSITY

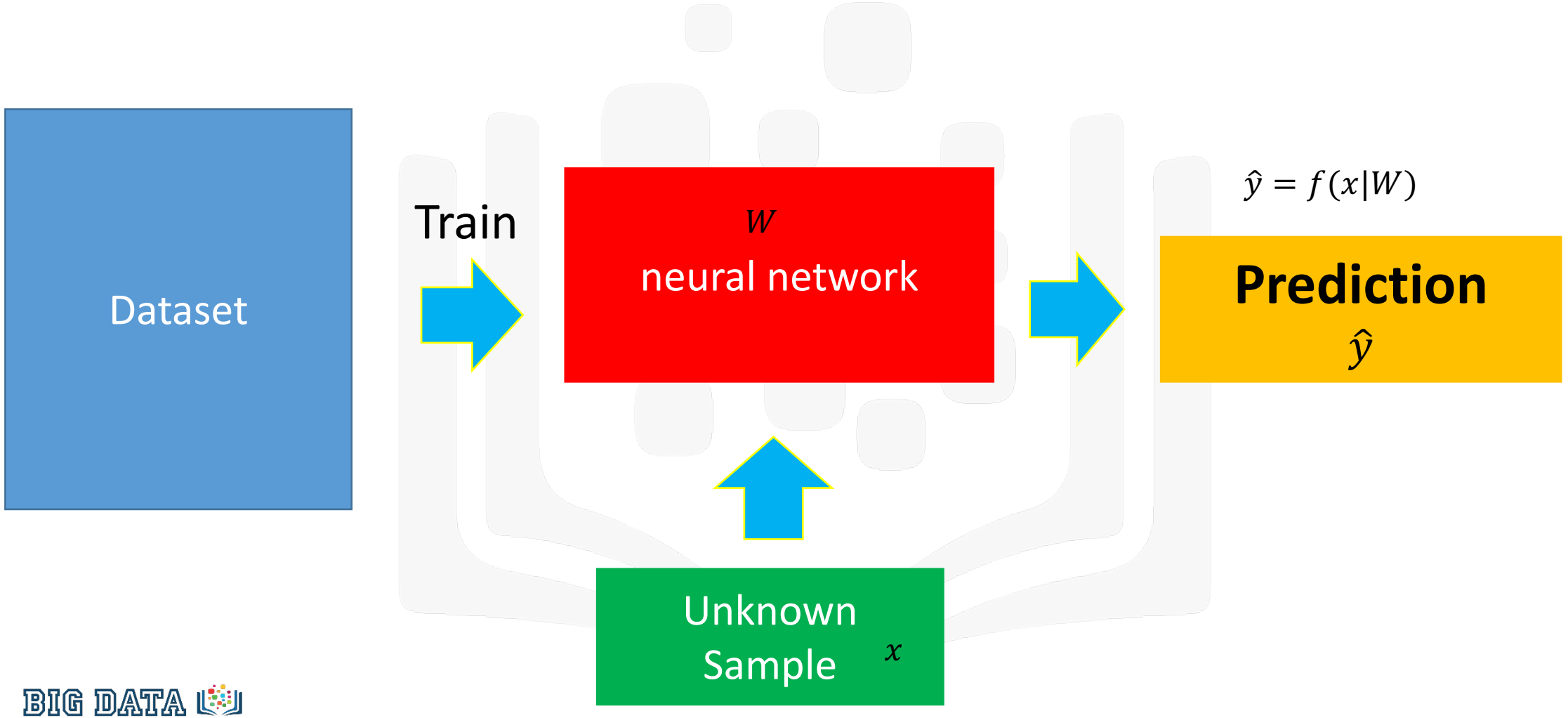
Neural Networks and Classification

- There a many applications of Neural Networks

1. Regression
2. Clustering
3. Recommendation Systems

- We will only Cover Classification

The Big Picture



BIG DATA UNIVERSITY

Targets :*Iris* Flower Data Set

- Targets are denoted as y
- For example $y \in \{0,1,2\}$ means y can equal: 0, 1, 2

$y = 0$



$y = 1$



$y = 2$



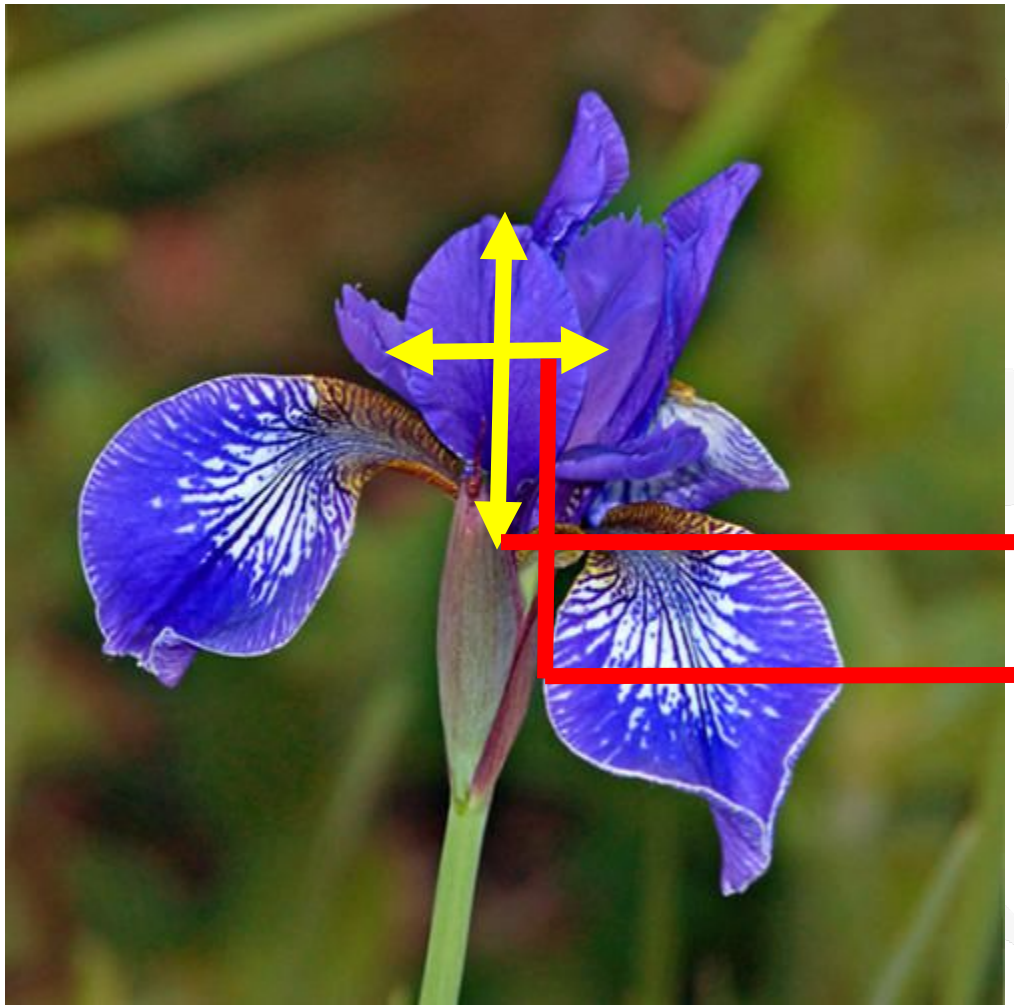
Features : Flower Data Set



Features				
Sepal length	Sepal width	Petal length	Petal width	Species
$(x)_1$	$(x)_2$	$(x)_3$	$(x)_4$	y

BIG DATA UNIVERSITY www.pinterest.com/aussiels/irises/

Iris Flower Data Set



Fisher's Iris Data				
Sepal length	Sepal width	Petal length	Petal width	Species
$(x)_1$	$(x)_2$	$(x)_3$	$(x)_4$	y

BIG DATA UNIVERSITY www.pinterest.com/aussiels/irises/

Training Set



$y_1 = 0$

$\mathbf{x}_1 = [4.9,3,1.4,0.2]$



$y_2 = 1$

$\mathbf{x}_2 = [4.2,2,1.4,0.3]$



$y_3 = 0$

$\mathbf{x}_3 = [4.5,2.9,1.3,0.4]$



$y_5 = 1$

$\mathbf{x}_5 = [4.5,2.1.3,0.4]$



$y_4 = 1$

$\mathbf{x}_4 = [3.9,2,3.4,0.6]$



$y_6 = 1$

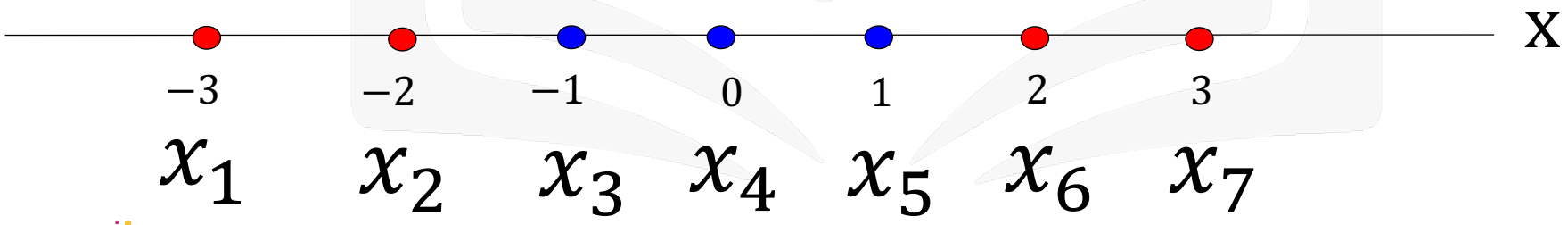
$\mathbf{x}_6 = [3.9,2,3.4,0.6]$

BIG DATA UNIVERSITY

Features and targets: Example

- Colors are used to indicates the class

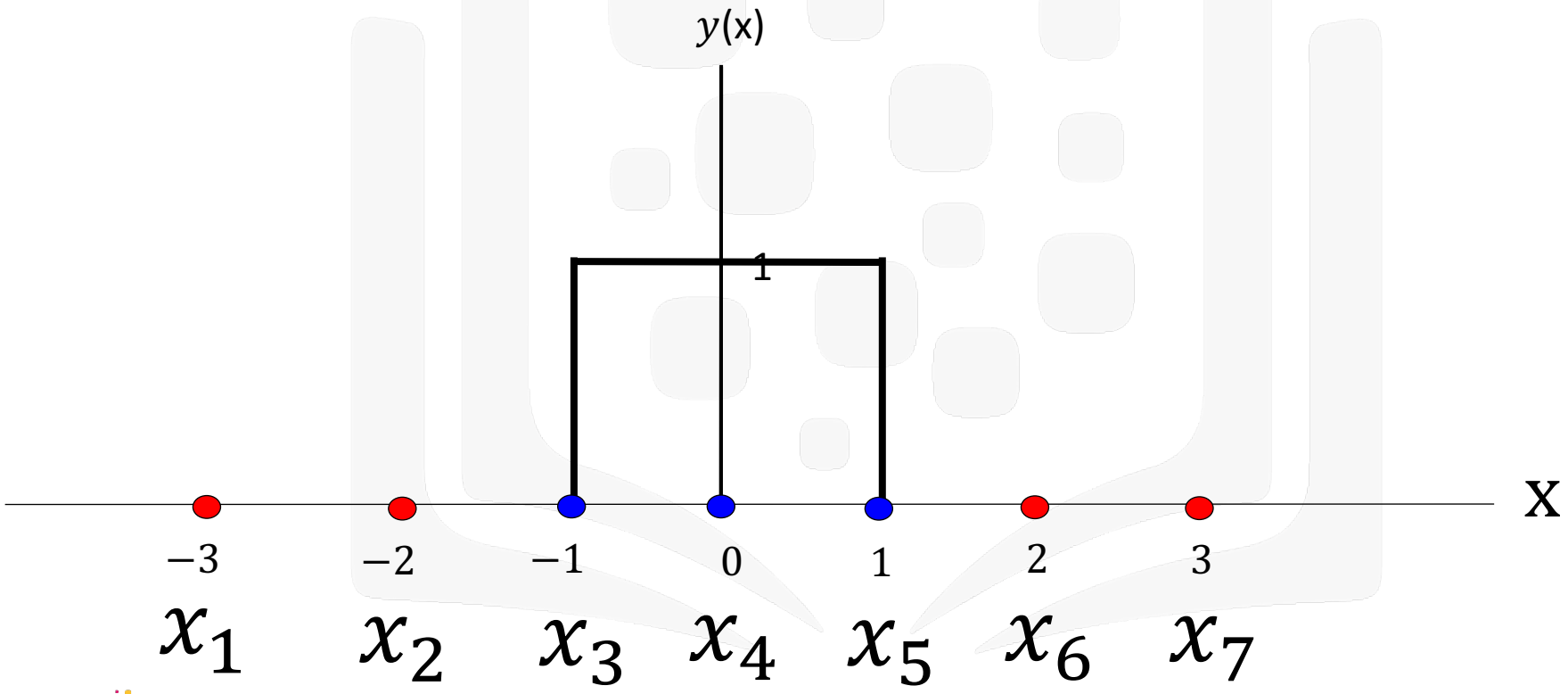
1. $y_1 = 0, x_1 = -3$
2. $y_2 = 0, x_2 = -2$
3. $y_3 = 1, x_3 = -1$
4. $y_4 = 1, x_4 = -1$
5. $y_5 = 1, x_5 = 1$
6. $y_5 = 0, x_6 = 2$
7. $y_7 = 0, x_7 = 3$



© 2015 BigDataUniversity.com

Features and targets: Example

- It is helpful to view the sample y as a decision function of x

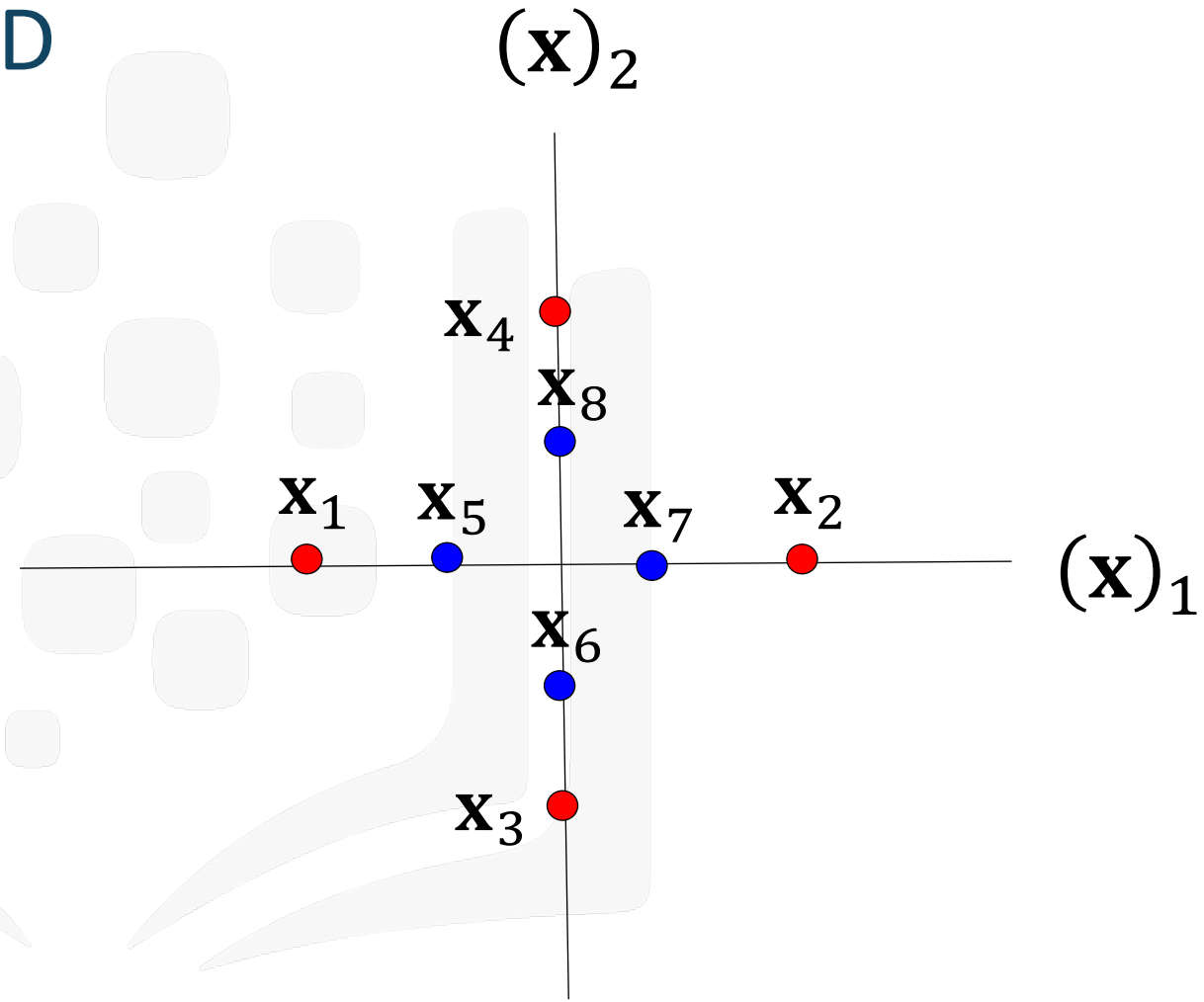


© 2015 BigDataUniversity.com

Features and targets: Example

- We can do the same in 2D

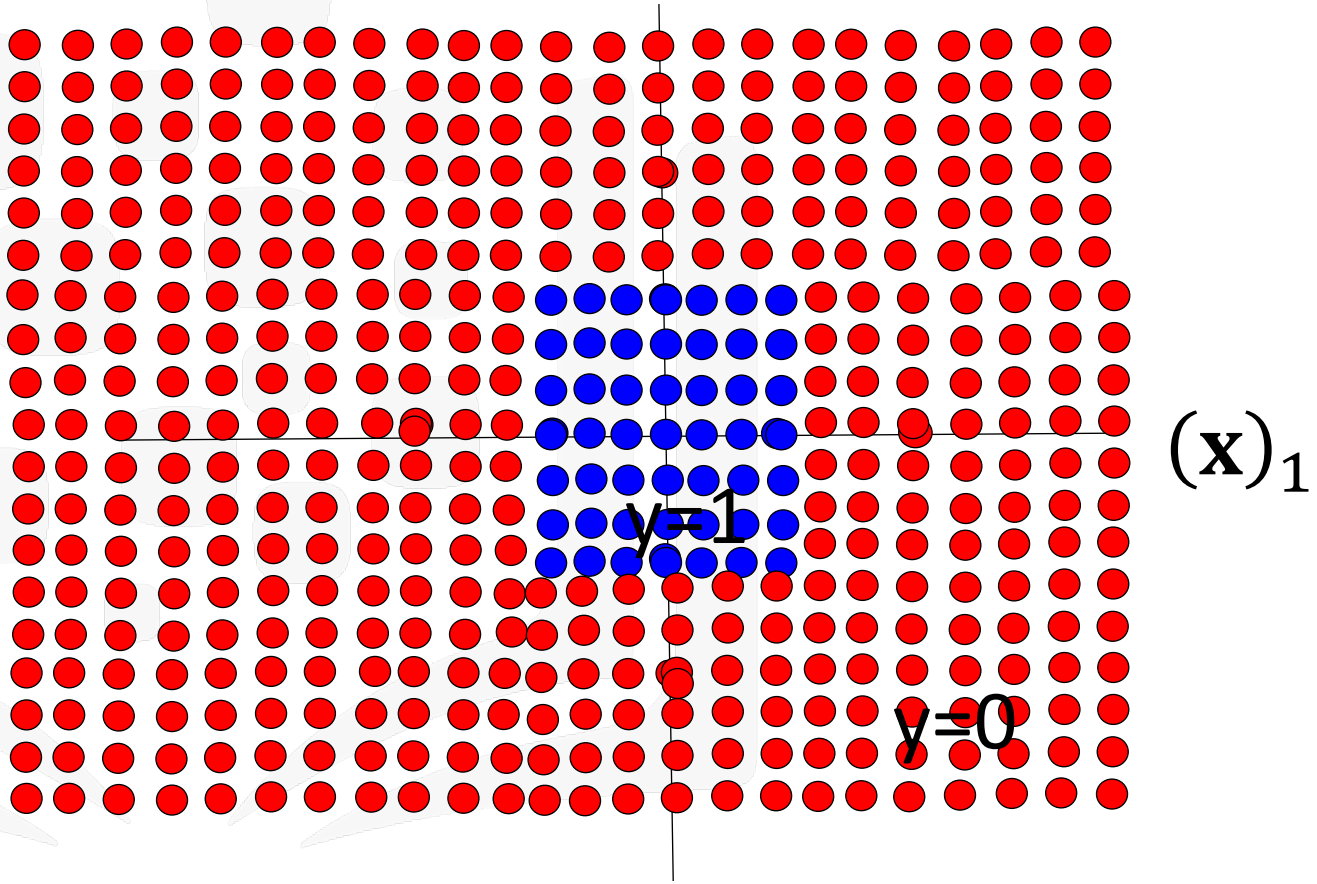
1. $y_1 = 0, \mathbf{x}_1 = [-2,\ 0]$
2. $y_2 = 0, \mathbf{x}_2 = [\ 2, 0]$
3. $y_3 = 0, \mathbf{x}_3 = [\ 0, -2]$
4. $y_4 = 0, \mathbf{x}_3 = [\ 0, 2]$
5. $y_5 = 1, \mathbf{x}_2 = [-1, 0]$
6. $y_6 = 1, \mathbf{x}_3 = [0, -1]$
7. $y_7 = 1, \mathbf{x}_1 = [1, 0]$
8. $y_8 = 1, \mathbf{x}_2 = [0, 1]$



© 2015 BigDataUniversity.com

Features and targets: Example

- We can also interpret these samples as as a function

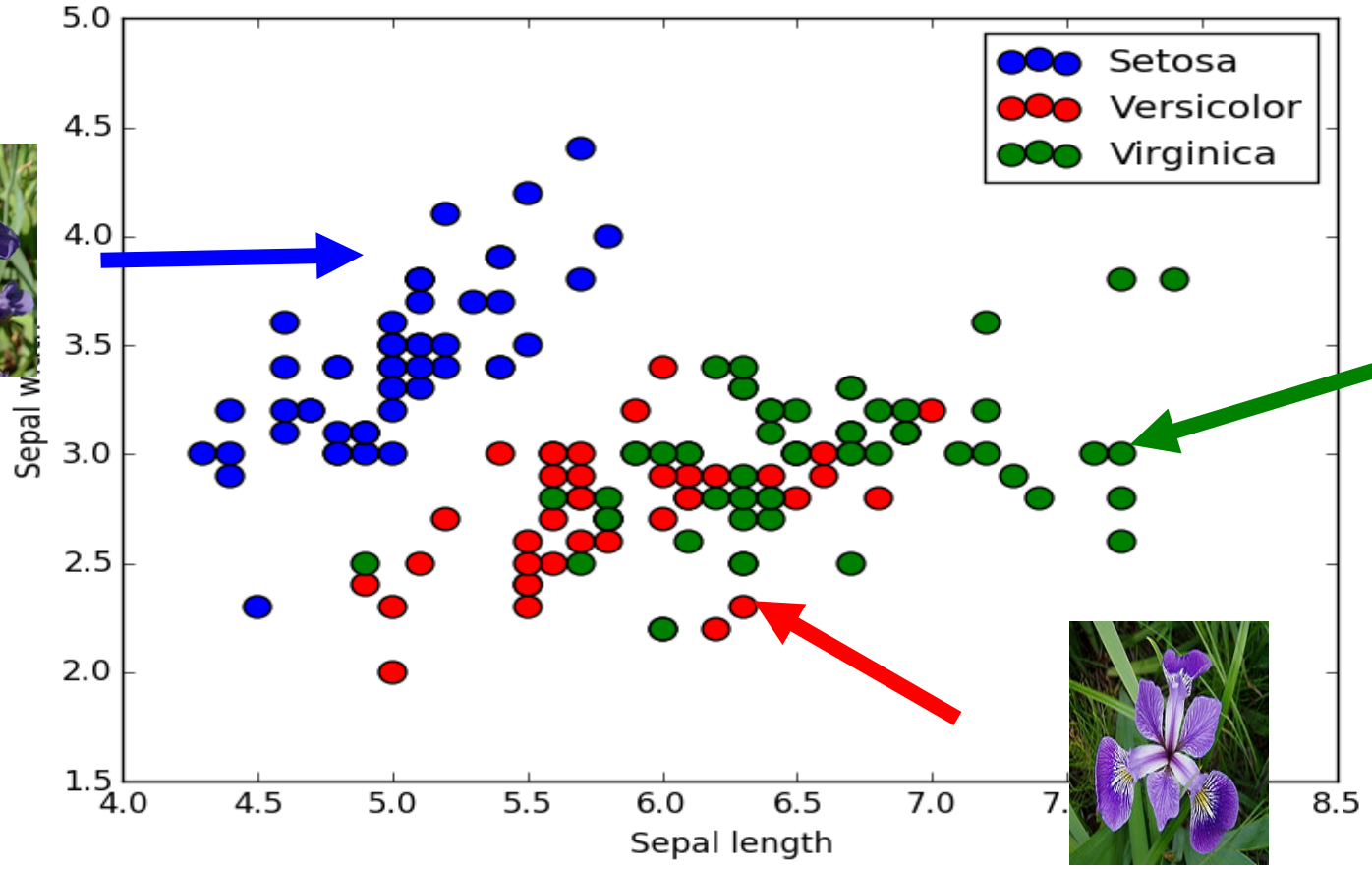


© 2015 BigDataUniversity.com

Features



0



1



2

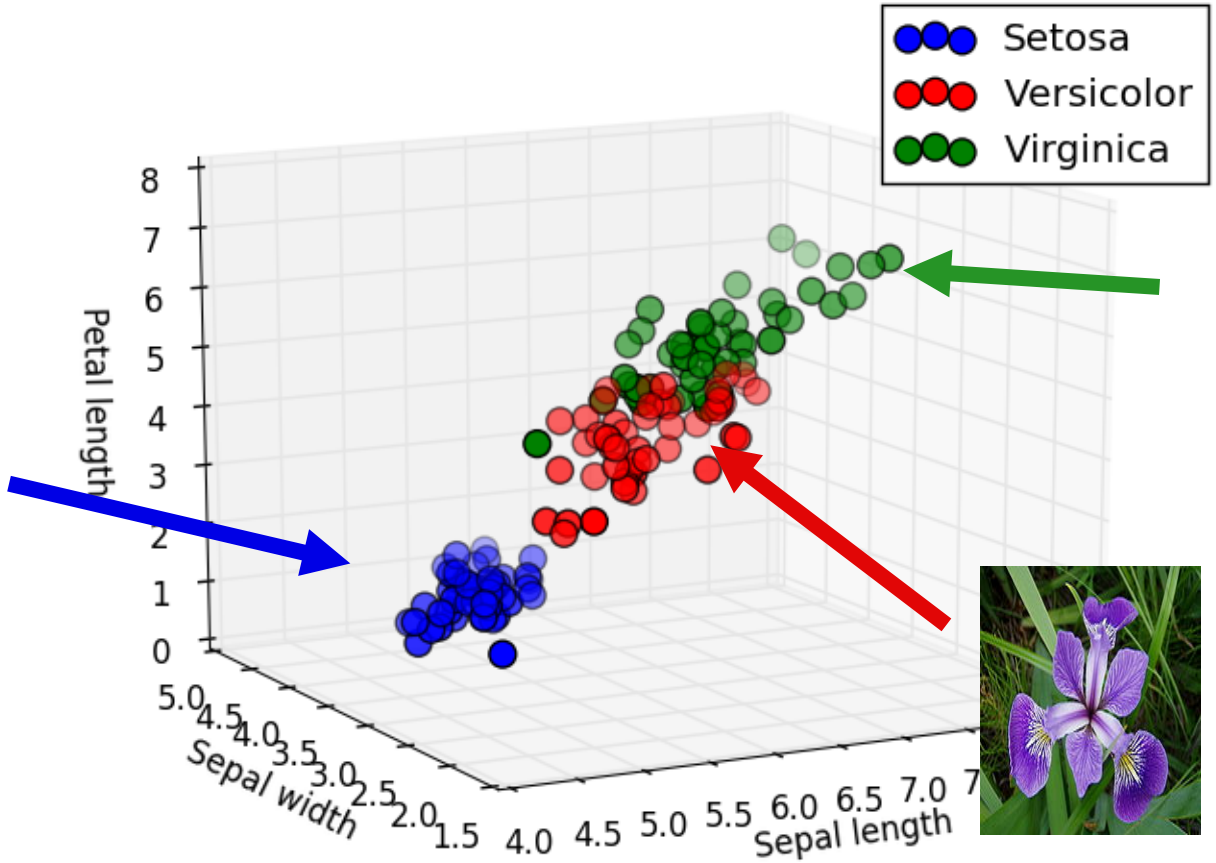
BIG DATA UNIVERSITY

Features

- Usually more features can give better performance



0



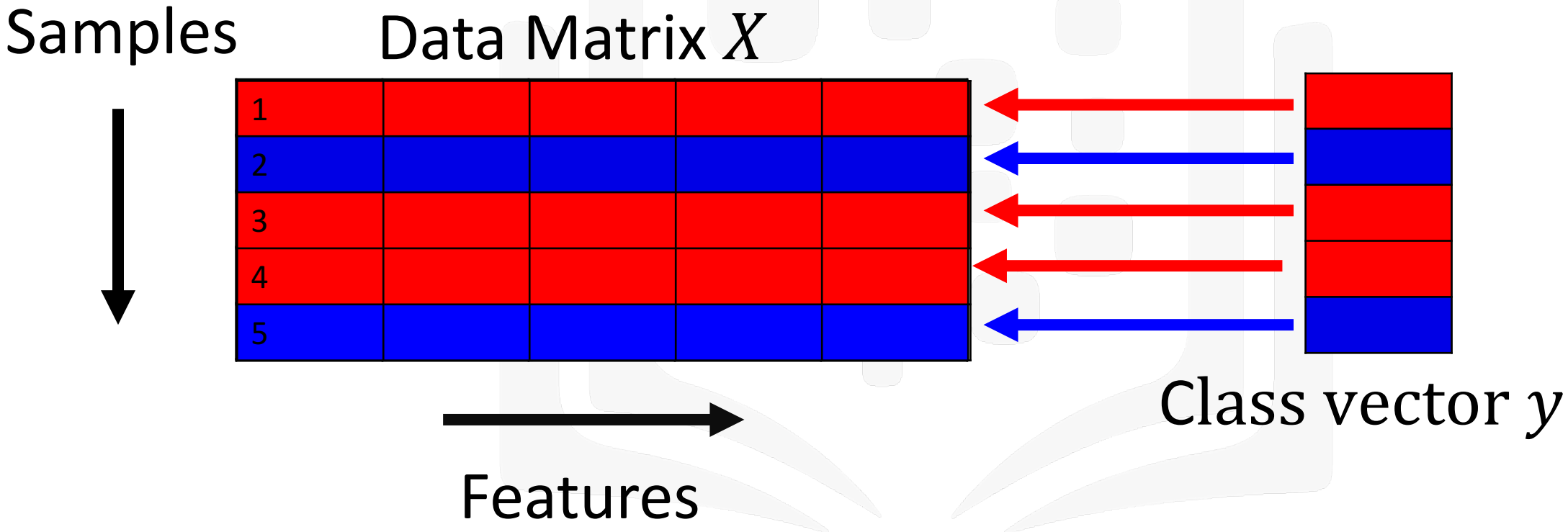


1

2

Features and Targets

- Data matrix and class vector



© 2015 BigDataUniversity.com

Features

- Color Images are comprised of different channels
- Each channel has an element represented with a number



71	61	96	122	117	80	72	74	79	76	87	88	75	112	78	72	76
122	119	94	48	64	61	69	73	76	62	49	101	109	98	86	78	74
66	55	49	37	42	54	64	63	79	56	105	70	83	66	97	84	77
58	49	50	66	78	79	77	59	43	54	26	66	102	69	97	87	83
93	80	110	109	86	71	68	114	101	92	86	64	34	88	109	100	102
78	76	97	92	88	45	118	98	96	95	73	44	63	84	92	110	102
59	79	73	71	139	126	85	79	65	93	105	62	80	74	86	101	92
82	84	126	114	155	51	96	65	30	73	82	76	158	58	88	90	88
71	76	56	62	132	41	80	94	25	78	26	51	72	90	86	80	82
73	69	54	24	162	92	89	48	68	88	72	85	155	41	93	74	73
72	100	55	28	75	54	75	103	81	15	102	89	218	89	81	99	81
89	42	91	33	43	70	70	112	110	121	150	228	57	69	112	119	101
89	77	54	47	41	50	87	31	83	125	84	53	57	109	93	126	101
55	36	35	36	37	54	64	18	126	39	54	86	38	73	94	94	78
63	33	25	29	43	72	59	29	21	37	22	46	89	53	98	78	71
68	53	46	34	56	74	80	117	48	52	29	50	99	26	74	85	101
74	69	44	48	69	75	82	123	64	56	22	53	20	37	73	90	93

22	26	44	62	54	37	32	26	31	29	47	40	39	38	20	33	36
48	51	40	26	35	37	30	26	29	30	26	42	38	29	18	31	35
22	20	35	23	21	25	27	24	32	28	217	28	31	8	19	35	33
22	20	22	29	31	32	36	29	19	162	38	69	222	18	24	18	31
33	30	47	49	36	38	32	234	231	251	212	192	34	35	49	32	35
41	38	41	41	42	16	210	221	221	239	209	186	29	36	28	38	36
34	42	32	83	212	64	171	197	194	218	189	155	22	30	23	29	31
34	39	253	250	254	87	81	190	128	194	176	188	214	17	24	22	37
28	218	226	177	193	115	37	36	70	146	111	128	158	109	46	47	35
28	223	231	161	253	211	44	26	70	75	77	100	188	23	53	32	38
32	205	230	207	193	24	37	87	67	13	179	169	254	224	49	51	42
37	22	176	15	13	28	39	100	108	231	255	239	211	237	236	57	52
28	31	31	22	16	26	51	24	63	208	255	223	255	253	35	52	55
26	22	26	20	18	17	34	14	93	23	61	200	201	212	44	29	32
24	19	17	16	19	34	26	17	16	21	21	20	196	30	61	40	36
25	26	27	21	20	30	22	81	30	34	22	23	149	17	33	24	62
27	31	22	18	25	22	24	75	37	22	15	18	12	19	32	34	43

69	70	110	132	119	80	64	53	56	59	71	67	76	126	80	65	70
122	121	97	53	70	71	61	52	54	57	49	69	107	103	90	75	73
52	48	62	41	46	47	51	49	59	52	66	75	118	81	108	87	75
55	50	37	52	57	58	61	52	41	24	74	78	69	72	105	91	87
83	70	64	68	62	56	53	81	69	86	58	47	55	83	103	98	107
77	70	65	68	64	44	81	62	61	67	51	35	77	92	92	111	108
69	89	67	67	118	84	60	49	32	58	63	39	95	74	81	105	101
79	97	113	107	176	53	83	41	3	44	48	48	161	65	78	95	102
75	59	46	59	138	31	48	67	1	56	32	43	44	91	105	99	98
80	49	38	22	164	56	102	80	59	79	74	79	146	48	111	83	93
64	85	33	22	77	80	93	99	74	27	99	101	243	66	112	112	98
67	36	70	60	61	80	80	113	107	99	140	253	58	54	86	131	118
93	58	49	56	56	60	106	50	61	108	76	63	63	104	105	123	115
47	37	36	55	50	59	62	25	101	53	65	68	37	49	98	97	91
42	30	26	36	59	79	46	15	18	28	52	60	64	74	106	87	82
45	64	53	41	69	79	77	145	32	44	47	64	88	45	87	93	114
53	71	61	62	81	84	76	159	59	42	47	64	30	50	80	97	104

BIG DATA UNIVERSITY

Features

69	70	110	132	119	80	64	53	56	59	71	67	76	126	80	65	70	
122	121	97	53	70	71	61	52	54	57	49	69	107	103	90	75	73	
52	48	62	41	46	47	51	49	59	52	66	75	118	81	108	87	75	
83	70	64	68	62	56	53	81	69	86	58	47	55	83	103	98	107	
33	30	47	49	36	38	32	234	231	251	212	192	34	35	49	32	35	
41	38	41	41	42	16	210	221	221	239	209	186	29	36	28	38	36	
34	42	32	83	212	64	171	197	194	218	189	155	22	30	23	29	31	
34	39	253	250	254	87	81	190	128	194	176	188	214	17	24	22	37	
28	218	226	177	193	115	37	36	70	146	111	128	158	109	46	47	35	
28	223	231	161	253	211	44	26	70	75	77	100	188	23	53	32	38	
32	205	230	207	193	24	37	87	67	13	179	169	254	224	49	51	42	
37	22	176	15	13	28	39	100	108	231	255	239	211	237	236	57	52	
28	31	31	22	16	26	51	24	63	208	255	223	255	253	35	52	55	
26	22	26	20	18	17	34	14	93	23	61	200	201	212	44	29	32	
24	19	17	16	19	34	26	17	16	21	21	20	196	30	61	40	36	
25	26	27	21	20	30	22	81	30	34	22	23	149	17	33	24	62	
27	31	22	18	25	22	24	75	37	22	15	18	12	19	32	34	43	

22	26	44	62	54	37	32	26	31	29	47	40	39	38	20	33	36	
48	51	40	26	35	37	30	26	29	30	26	42	38	29	18	31	35	
22	20	35	23	21	25	27	24	32	28	217	28	31	8	19	35	33	
22	20	22	29	31	32	36	29	19	162	38	69	222	18	24	18	31	
33	30	47	49	36	38	32	234	231	251	212	192	34	35	49	32	35	
41	38	41	41	42	16	210	221	221	239	209	186	29	36	28	38	36	
34	42	32	83	212	64	171	197	194	218	189	155	22	30	23	29	31	
34	39	253	250	254	87	81	190	128	194	176	188	214	17	24	22	37	
28	218	226	177	193	115	37	36	70	146	111	128	158	109	46	47	35	
28	223	231	161	253	211	44	26	70	75	77	100	188	23	53	32	38	
32	205	230	207	193	24	37	87	67	13	179	169	254	224	49	51	42	
37	22	176	15	13	28	39	100	108	231	255	239	211	237	236	57	52	
28	31	31	22	16	26	51	24	63	208	255	223	255	253	35	52	55	
26	22	26	20	18	17	34	14	93	23	61	200	201	212	44	29	32	
24	19	17	16	19	34	26	17	16	21	21	20	196	30	61	40	36	
25	26	27	21	20	30	22	81	30	34	22	23	149	17	33	24	62	
27	31	22	18	25	22	24	75	37	22	15	18	12	19	32	34	43	

22	26	44	62	54	37	32	26	31	29	47	40	39	38	20	33	36	
48	51	40	26	35	37	30	26	29	30	26	42	38	29	18	31	35	
22	20	35	23	21	25	27	24	32	28	217	28	31	8	19	35	33	
22	20	22	29	31	32	36	29	19	162	38	69	222	18	24	18	31	
33	30	47	49	36	38	32	234	231	251	212	192	34	35	49	32	35	
41	38	41	41	42	16	210	221	221	239	209	186	29	36	28	38	36	
34	42	32	83	212	64	171	197	194	218	189	155	22	30	23	29	31	
34	39	253	250	254	87	81	190	128	194	176	188	214	17	24	22	37	
28	218	226	177	193	115	37	36	70	146	111	128	158	109	46	47	35	
28	223	231	161	253	211	44	26	70	75	77	100	188	23	53	32	38	
32	205	230	207	193	24	37	87	67	13	179	169	254	224	49	51	42	
37	22	176	15	13	28	39	100	108	231	255	239	211	237	236	57	52	
28	31	31	22	16	26	51	24	63	208	255	223	255	253	35	52	55	
26	22	26	20	18	17	34	14	93	23	61	200	201	212	44	29	32	
24	19	17	16	19	34	26	17	16	21	21	20	196	30	61	40	36	
25	26	27	21	20	30	22	81	30	34	22	23	149	17	33	24	62	
27	69	70	110	132	119	80	64	53	56	59	71	67	76	126	80	65	70

$$\mathbf{x}_1 = [\quad \ldots \ldots \quad]$$

BIG DATA UNIVERSITY

Linear Classifiers

BIG DATA UNIVERSITY

Linear Classifier

- The equation of a line in 1 dimension is given by:

$$wx + b$$

- This generalizes in D dimensions to :

$$\boldsymbol{w}^{\mathrm{T}}\mathbf{x}$$

- Lets see what happens for different values of $\mathbf{x}$ *?*

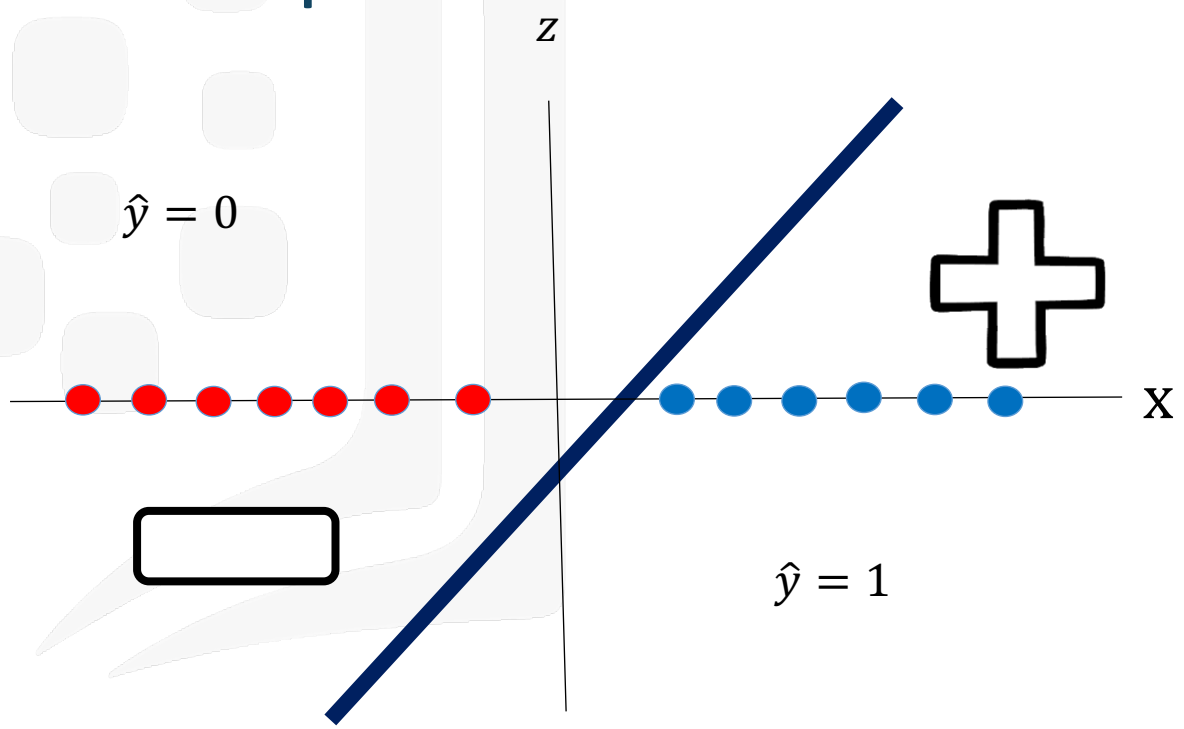
© 2015 BigDataUniversity.com

Linear Classifier

- Consider the following data set
- If we can separate the data with a line, we can use that line to classify the sample

$$z = wx + b$$

$$= 1x - 1$$



BIG DATA UNIVERSITY

© 2015 BigDataUniversity.com

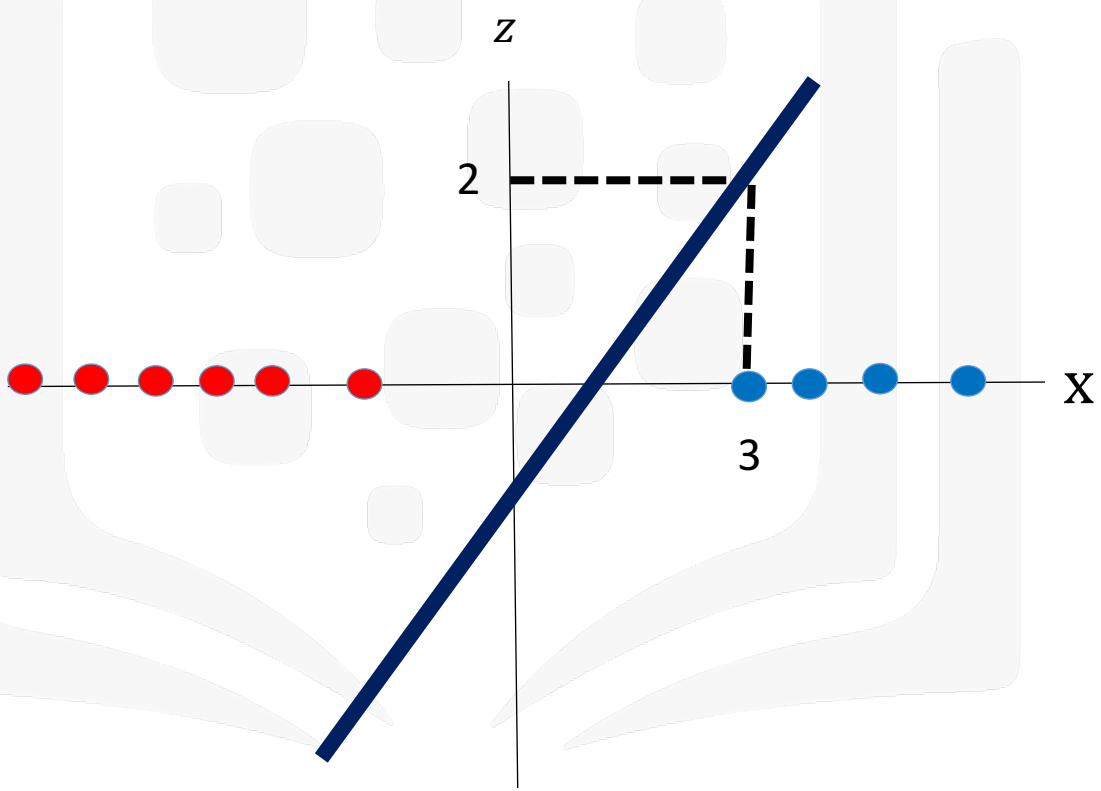
Linear Classifier

- If x is on the left side of the line we get a positive number

$$z = 1x - 1$$

$$z = 1(3) - 1$$

$$= 2$$



BIG DATA UNIVERSITY

© 2015 BigDataUniversity.com

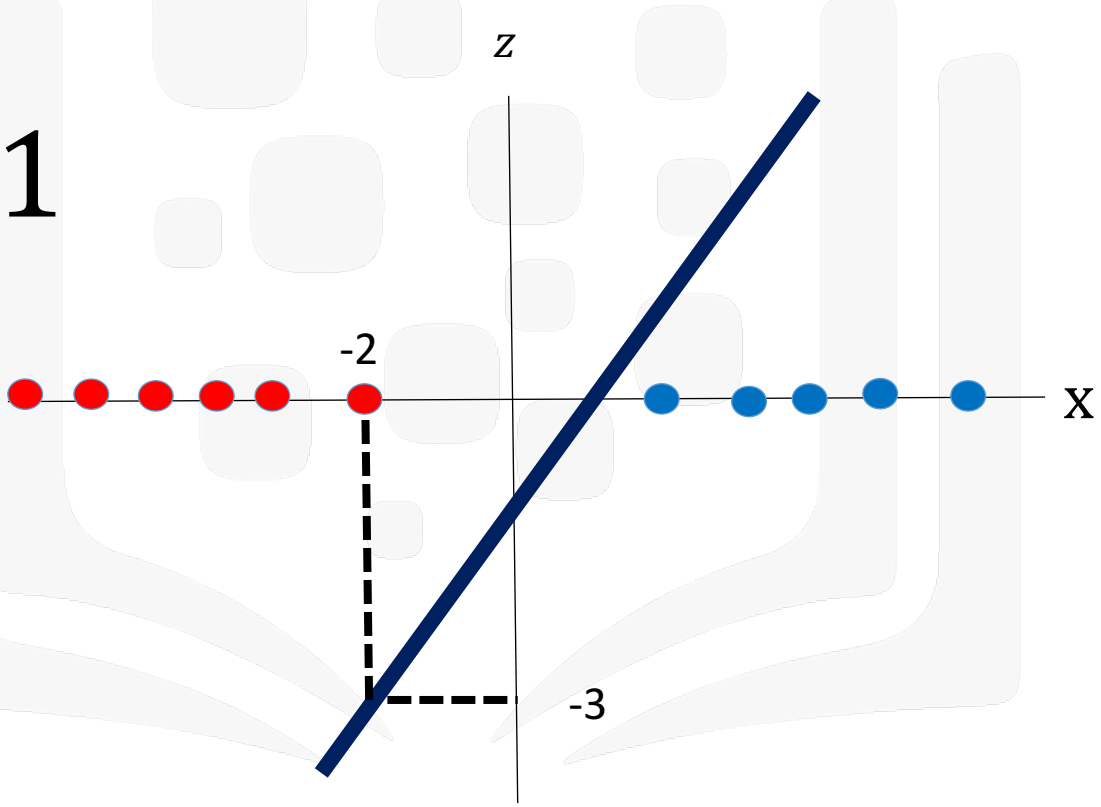
Linear Classifier

- If x is on the right side of the line we get a positive number

$$z = 1x - 1$$

$$z = 1(-2) - 1$$

$$= -3$$

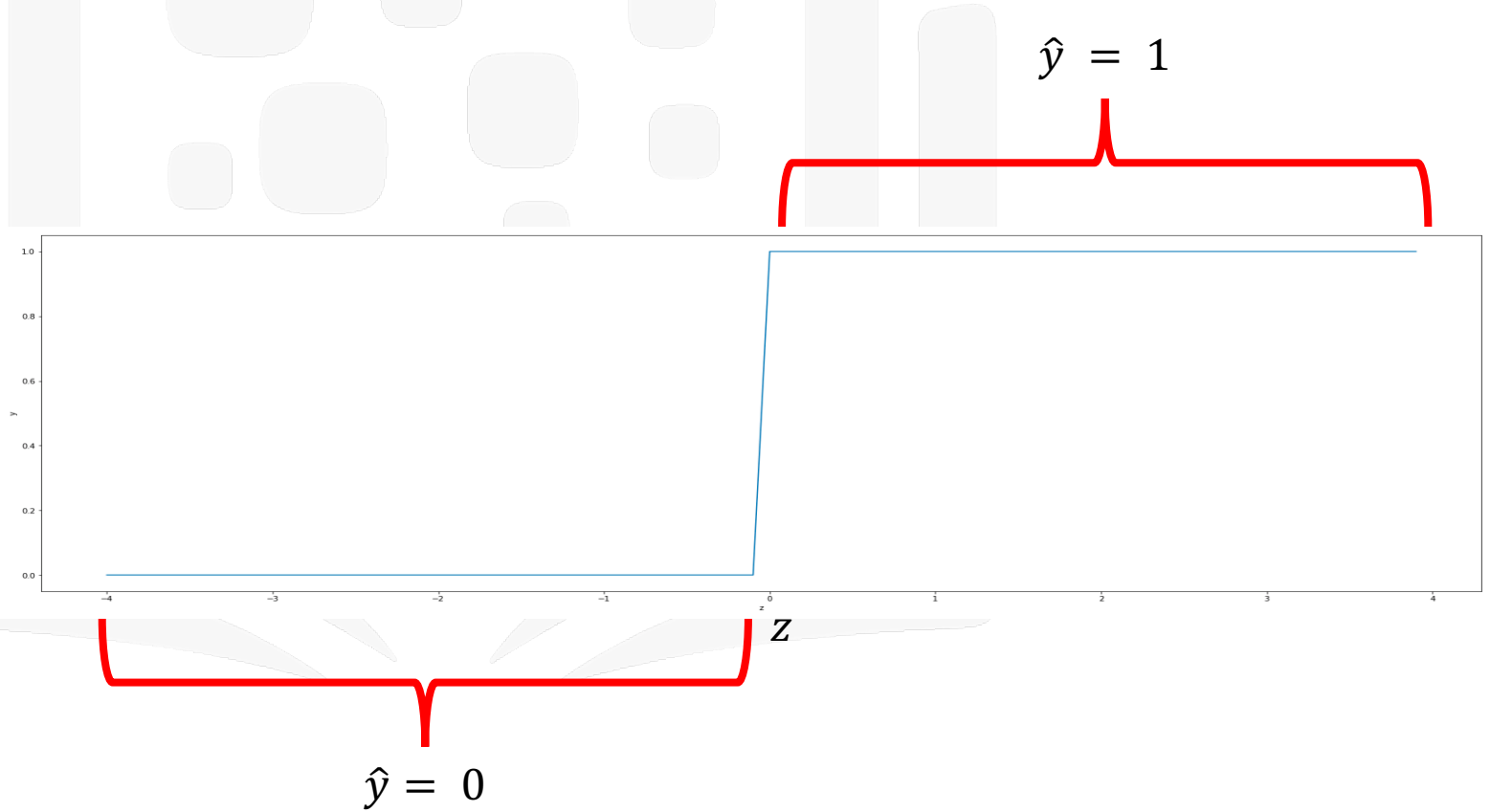


BIG DATA UNIVERSITY

© 2015 BigDataUniversity.com

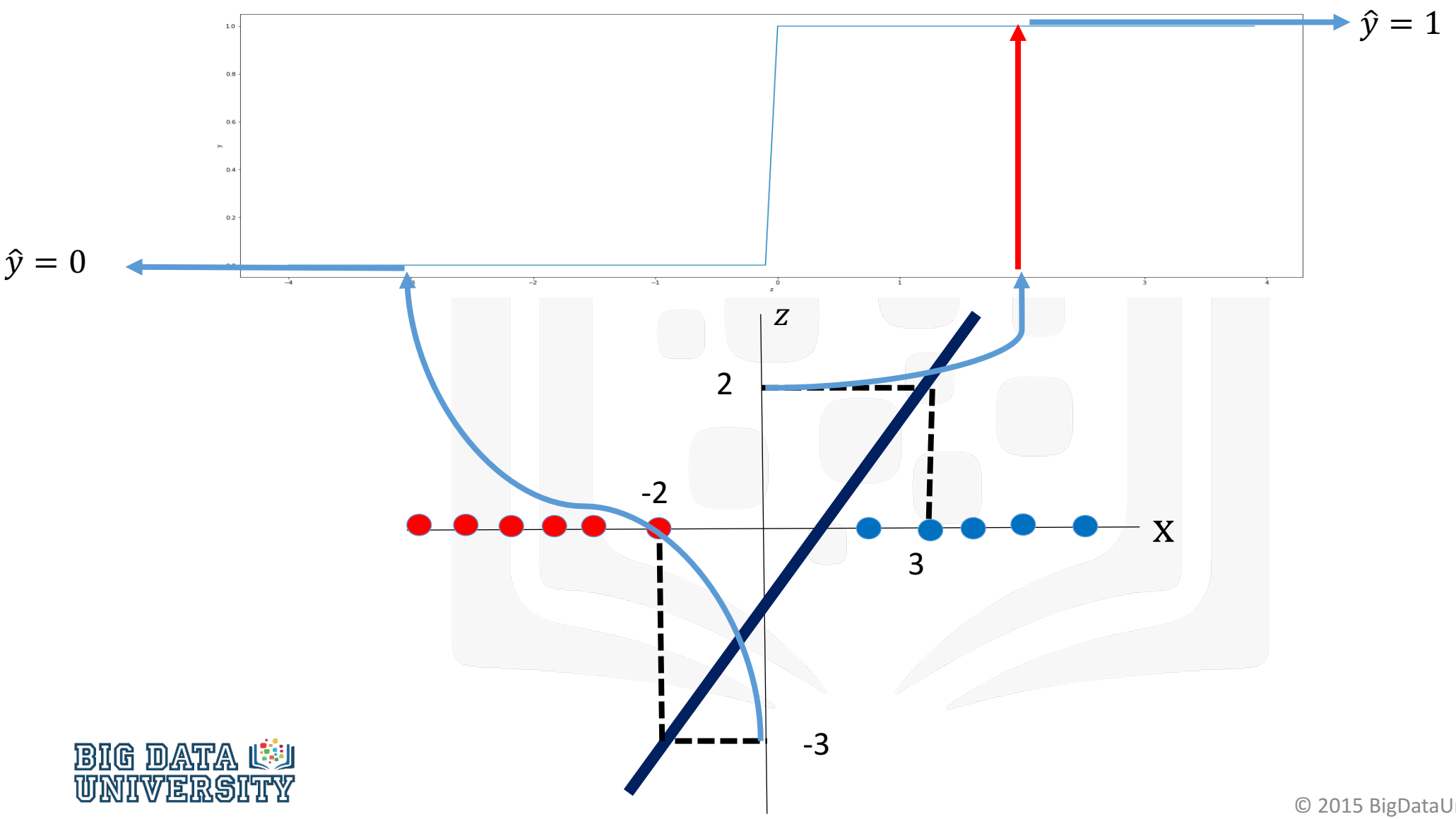
Linear Classifiers

- If we use the line to calculate the class of a point, it always returns a positive or negative numbers, such as 3, -2, and so on.
- But, we need class 0 and class 1. How we can convert the numbers into 0 and 1?
- $\hat{y} = 1$, if z>= 0
- $\hat{y} = 0$, if z< 0



BIG DATA UNIVERSITY

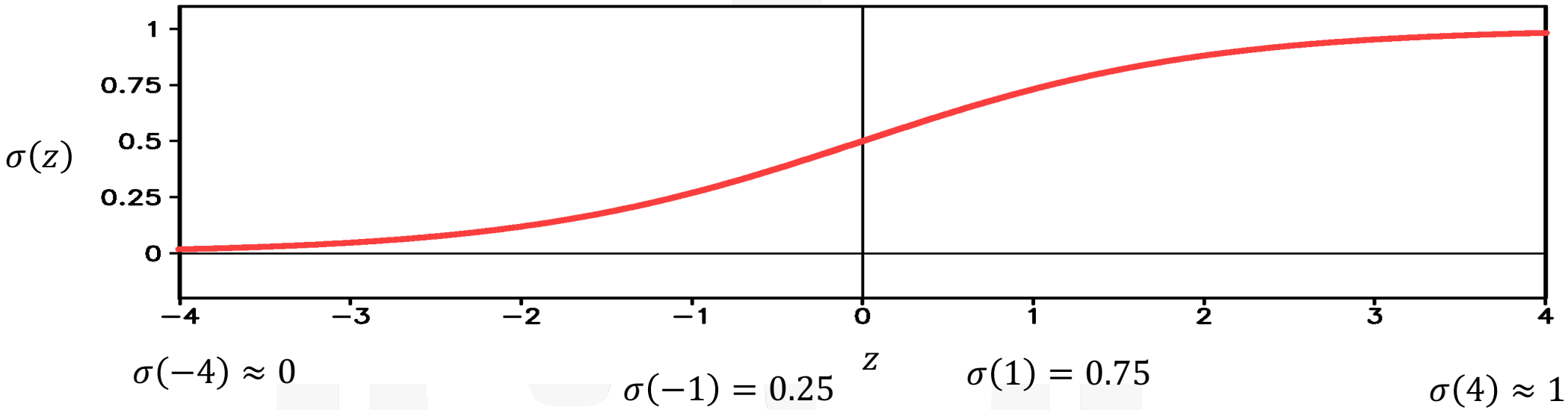
Linear Classifier: Threshold Function



© 2015 BigDataUniversity.com

Logistic regression

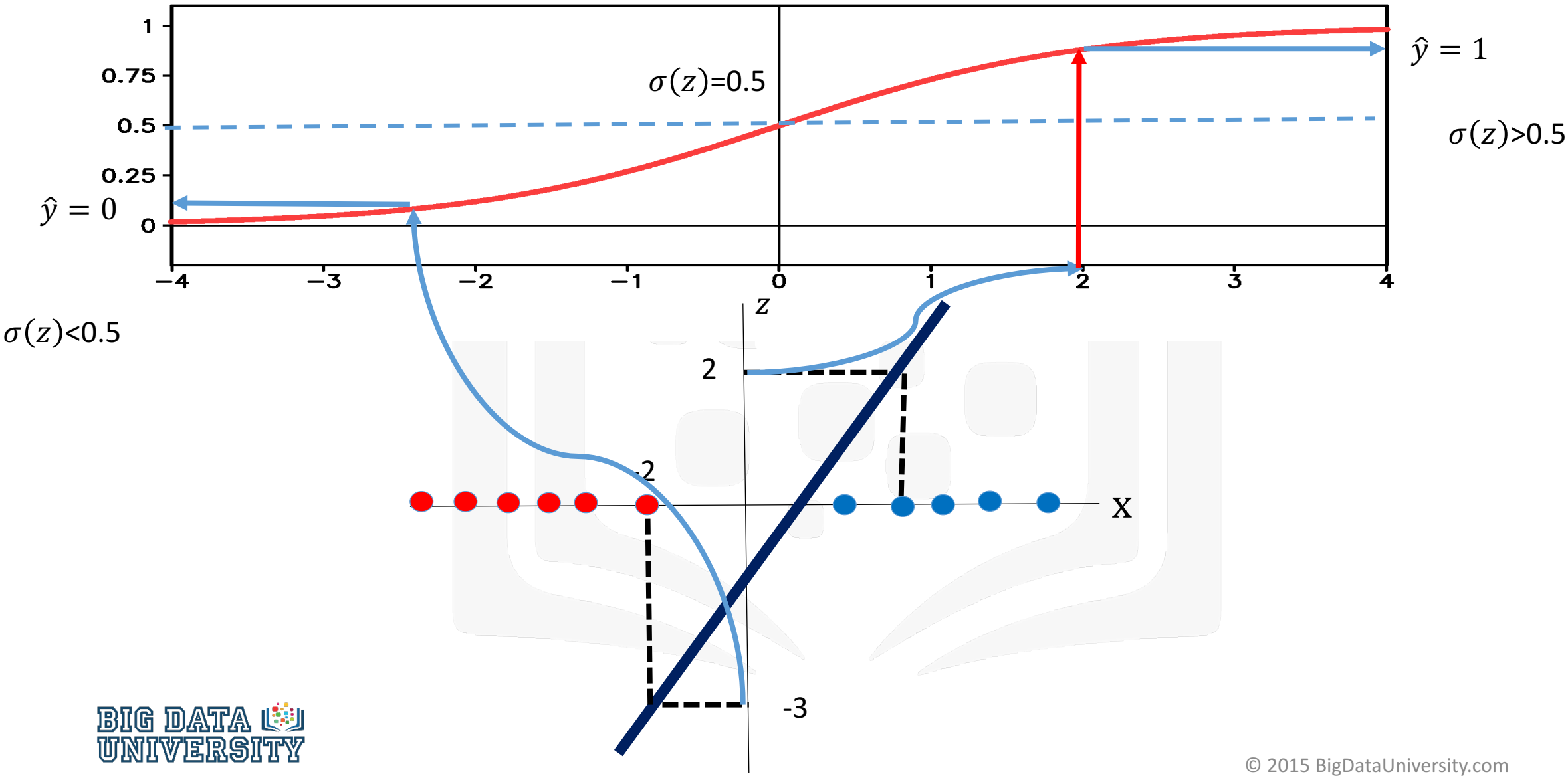
- Logistic Function



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

© 2015 BigDataUniversity.com

Logistic regression



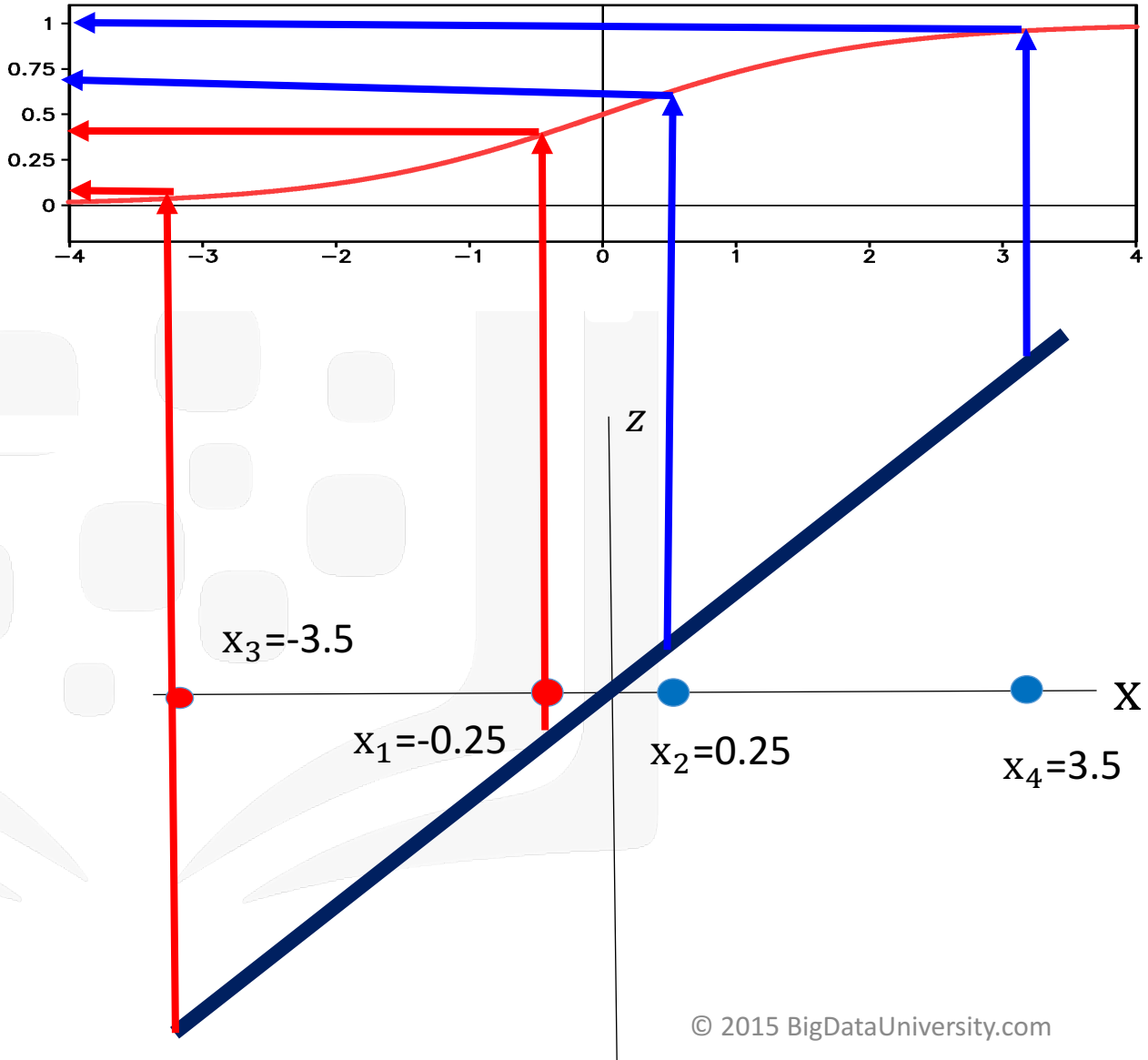
BIG DATA UNIVERSITY

© 2015 BigDataUniversity.com

Logistic regression

- Samples are near the line the bound we get values near 0.5
- If they are far we get values near 0 or 1

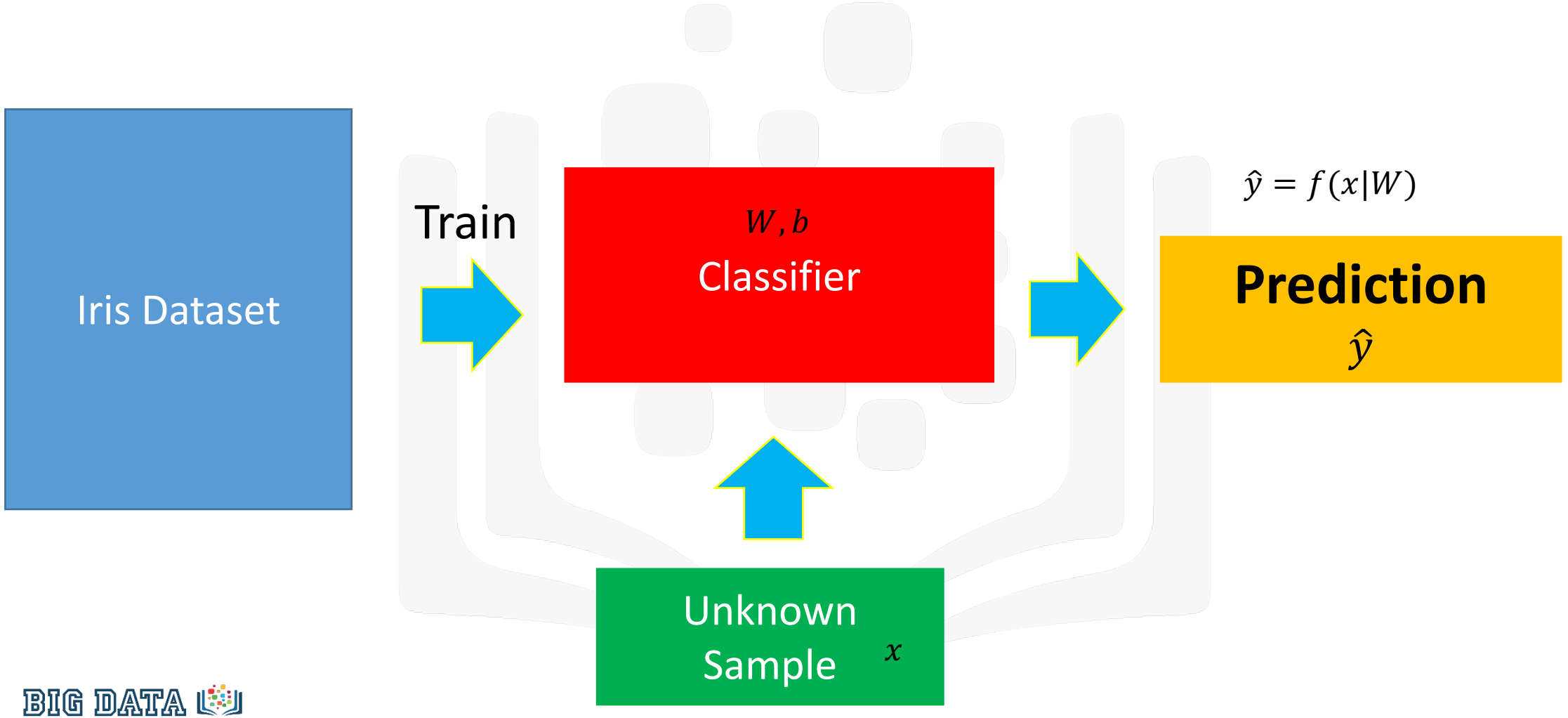
x	y	$\sigma(z)$	$\hat{y}$
x_1	0	0. 44	0
x_2	1	0.56	1
x_3	0	0. 02	0
x_4	1	0.98	1



BIG DATA UNIVERSITY

© 2015 BigDataUniversity.com

The Big Picture



BIG DATA UNIVERSITY

Training

BIG DATA UNIVERSITY

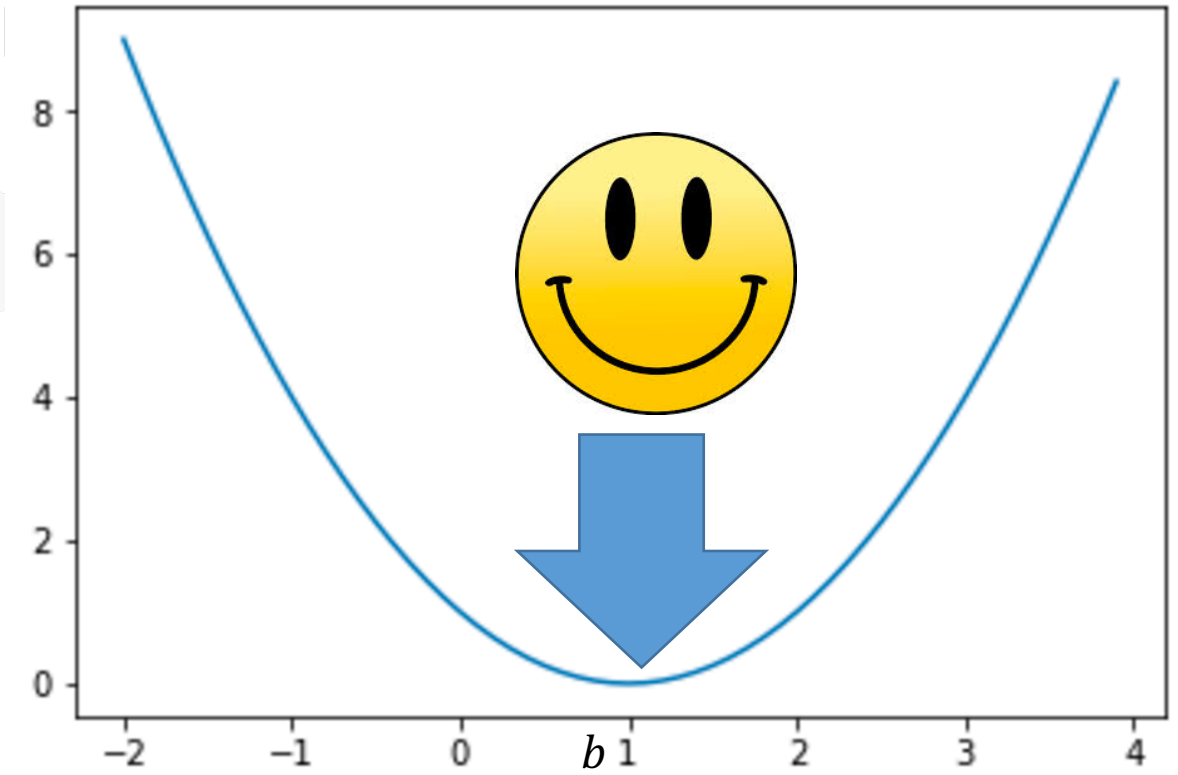
Training: Cost Function

- Lets try to train a model to predict $y = 1$
- Lets forget about the logistic function try a simple model: $\hat{y} = b$
- We use a cost function to see how good our model is :

$$(y - b)^2$$

$$C(b) = (1 - b)^2$$

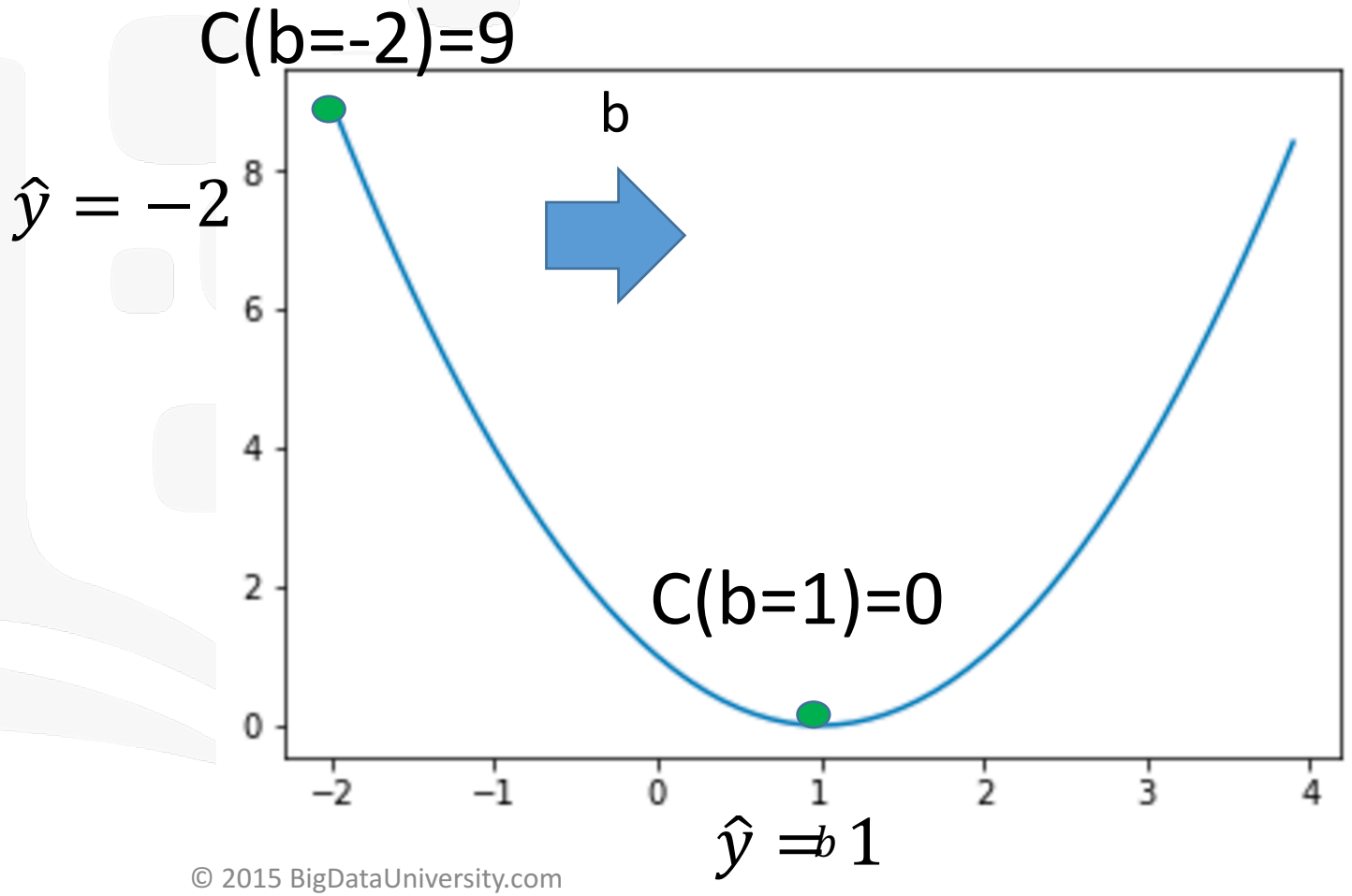
What's the best value for b ?



BIG DATA UNIVERSITY

Training

- Let say we started with a guess b=-2



© 2015 BigDataUniversity.com

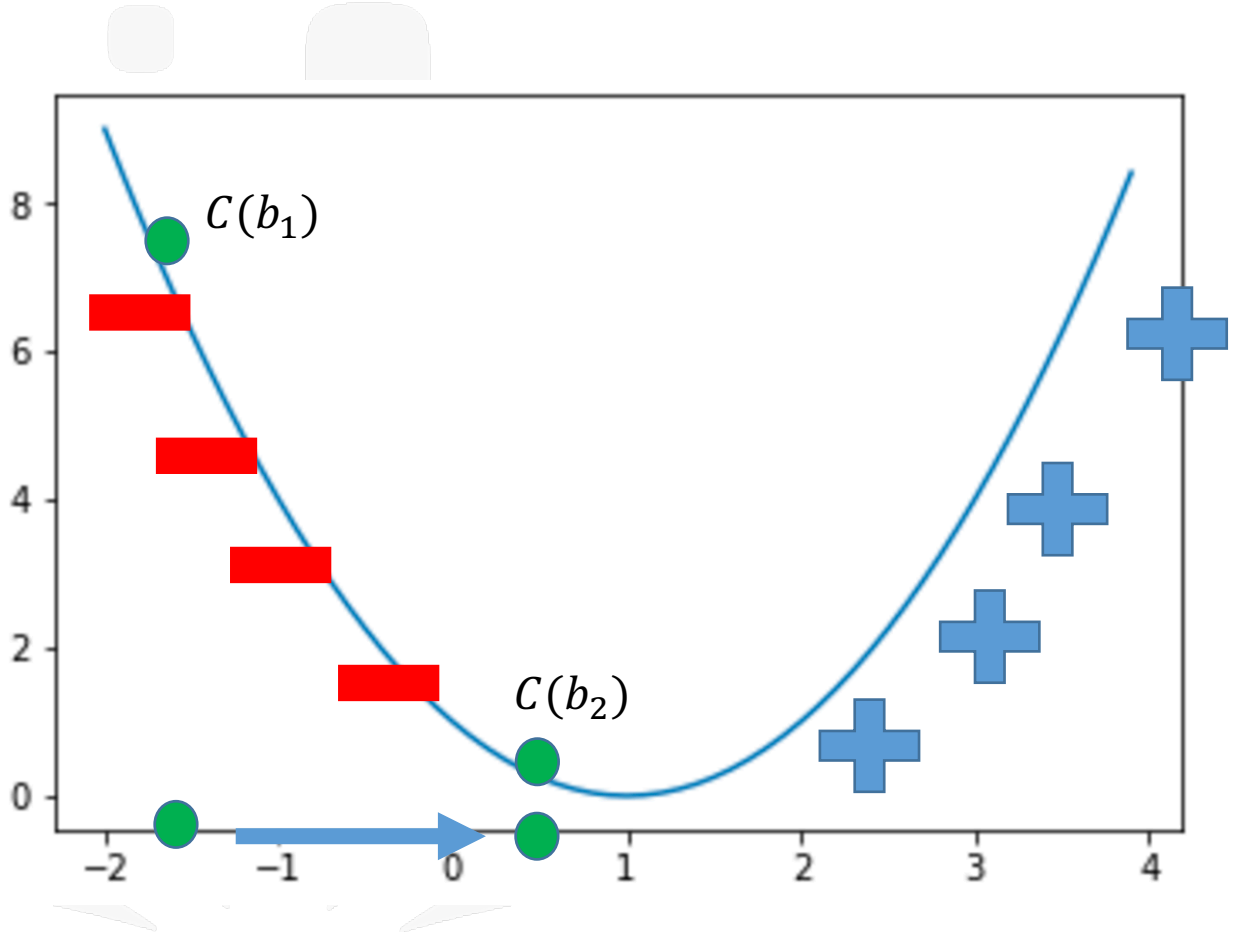
Training

$$slope\ (b)$$

Learning Rate

$$b_2 = b_1 - \eta\ slope(b_1)$$

$$b_2 = b_1 + number$$



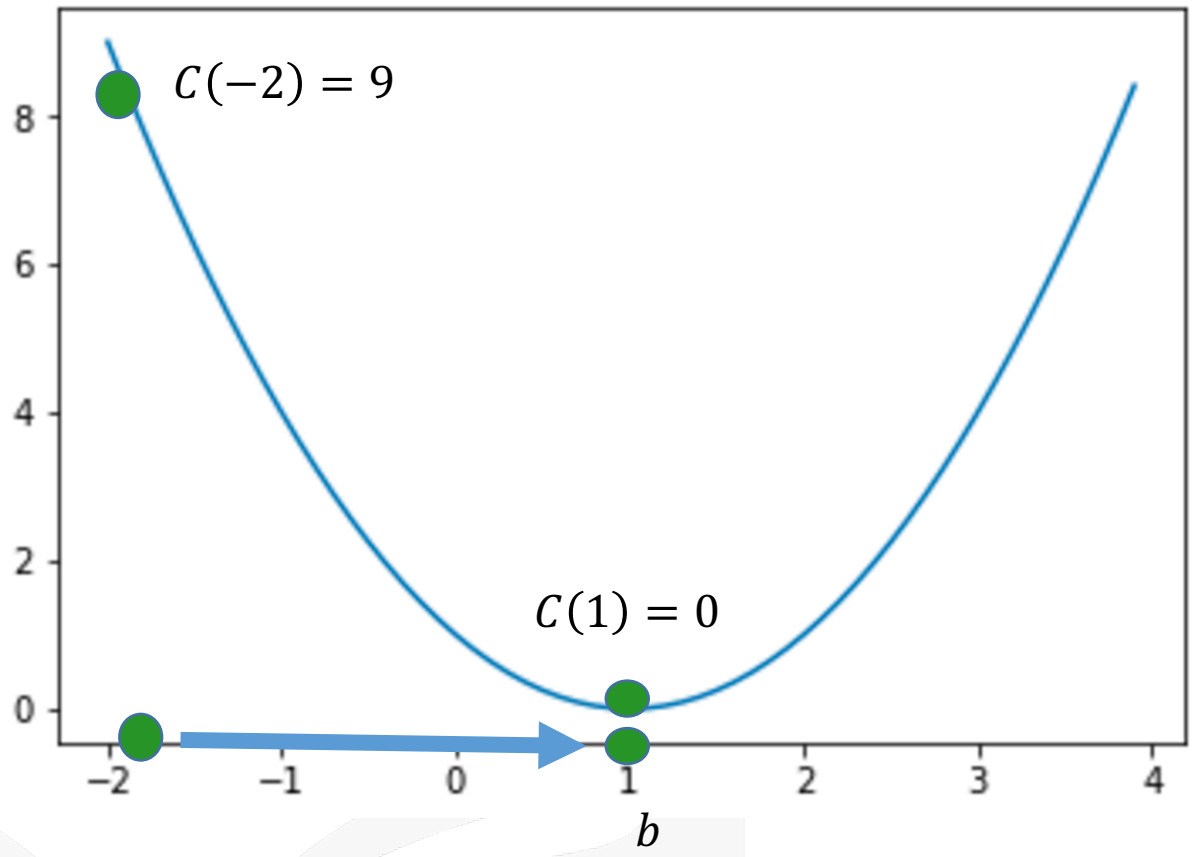
© 2015 BigDataUniversity.com

Training

$$b_2 = -2 - 1.5slope(-2)$$

$$b_1 = -2 + 3$$

$$b_1 = 1$$



$$\hat{y} = 1$$

© 2015 BigDataUniversity.com

Training: Cost

- All we do for the logistic regression just add the logistic function

$$c(w,b) = (y - b)^2$$

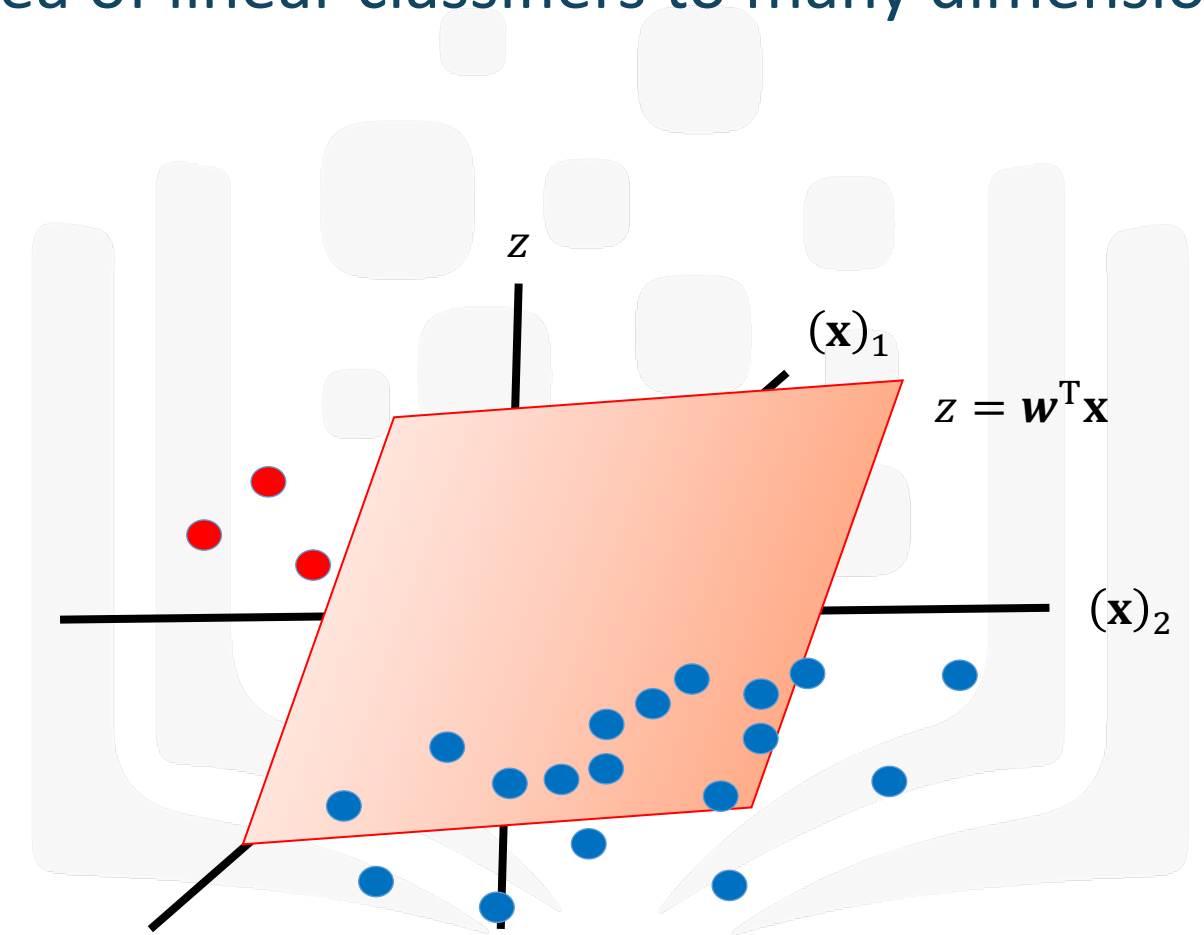
$$c(w,b) = (y - \sigma(wx + b))^2$$

$$c(w,b) = (y_1 - (wx_1 + b))^2 + .. + (y_N - (wx_N + b))^2$$

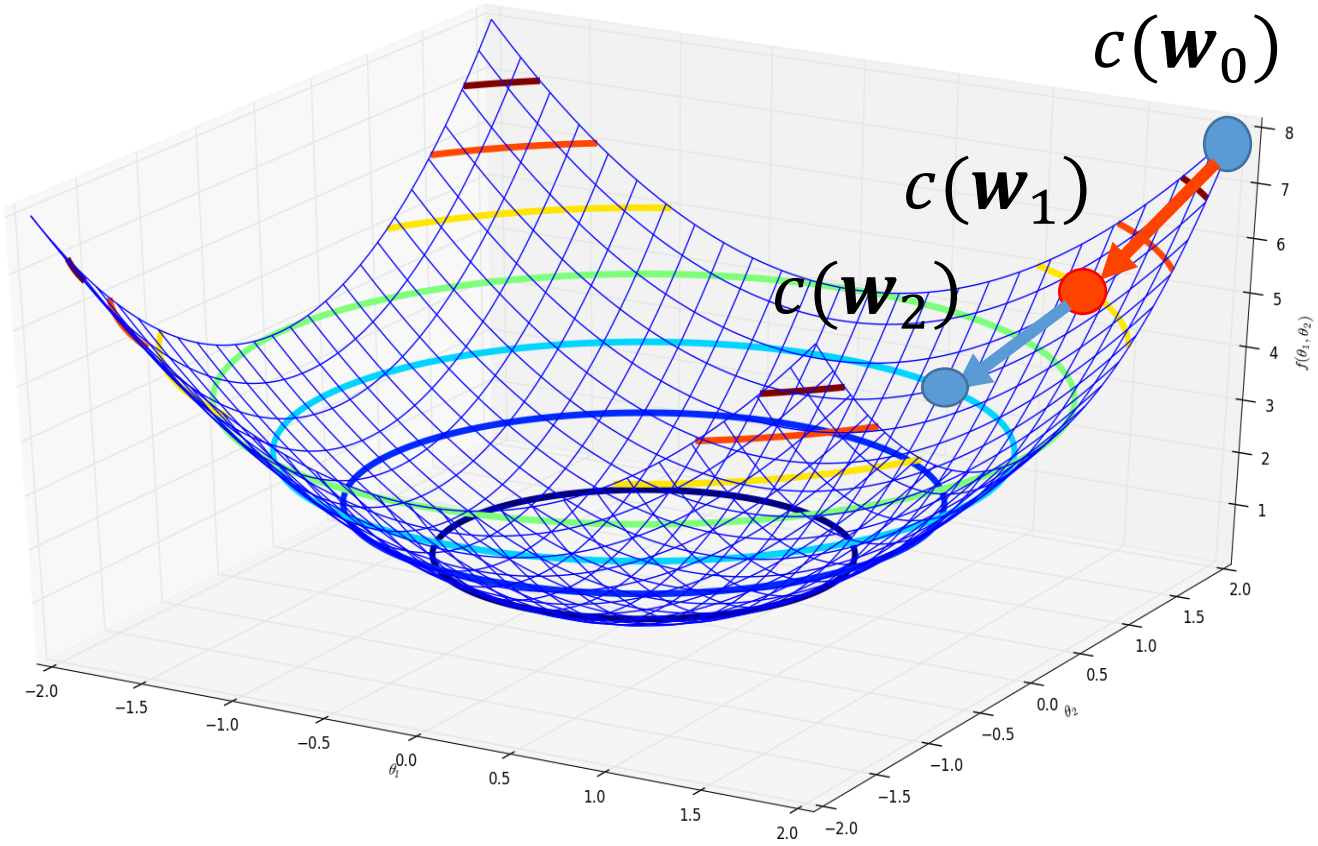
BIG DATA UNIVERSITY

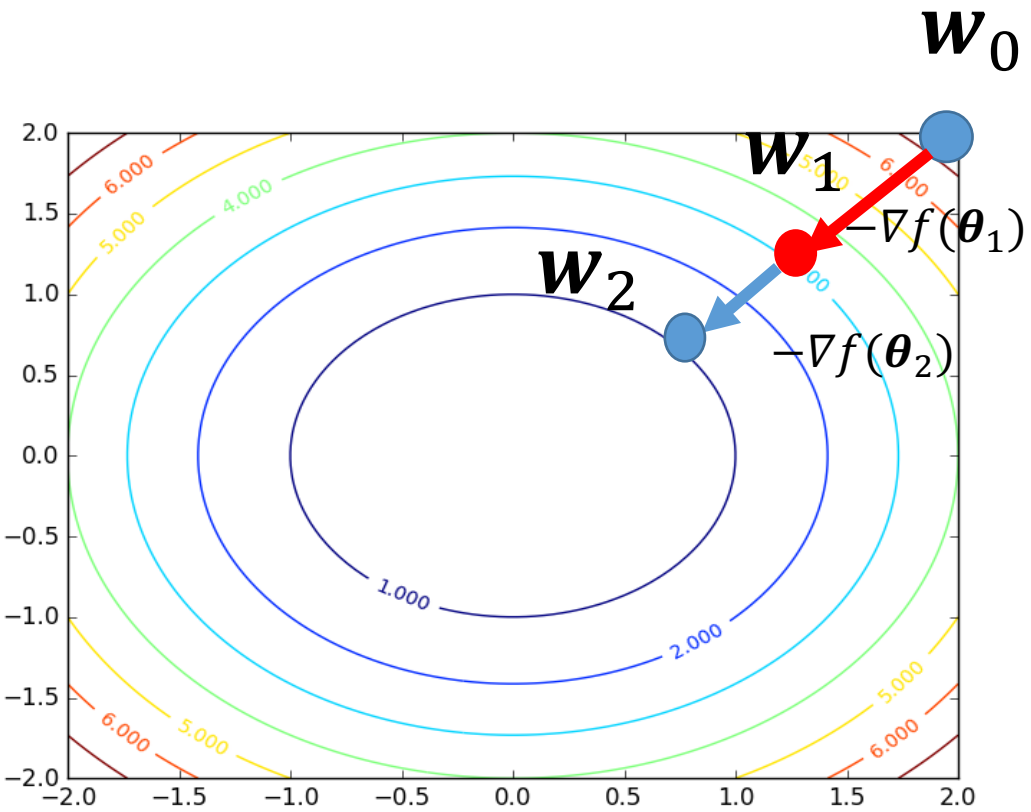
Linear Classifier

- The idea of linear classifiers to many dimensions

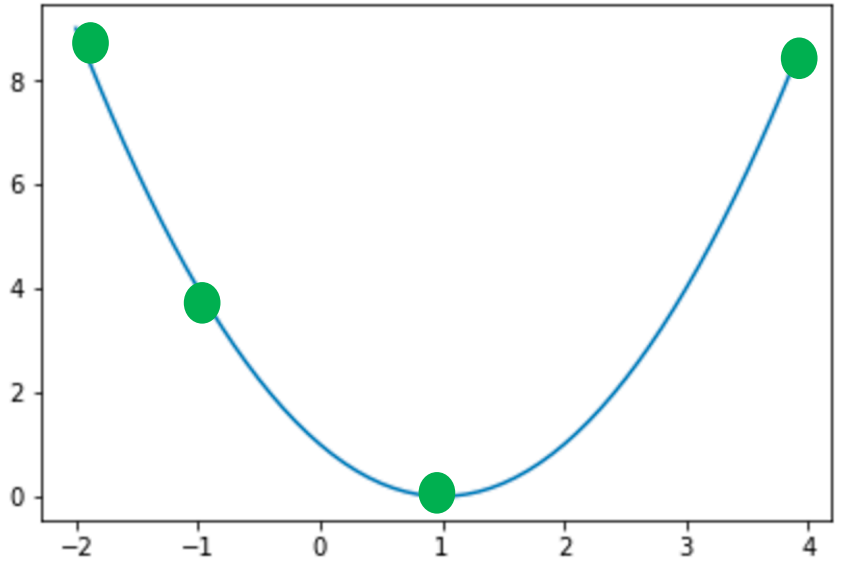


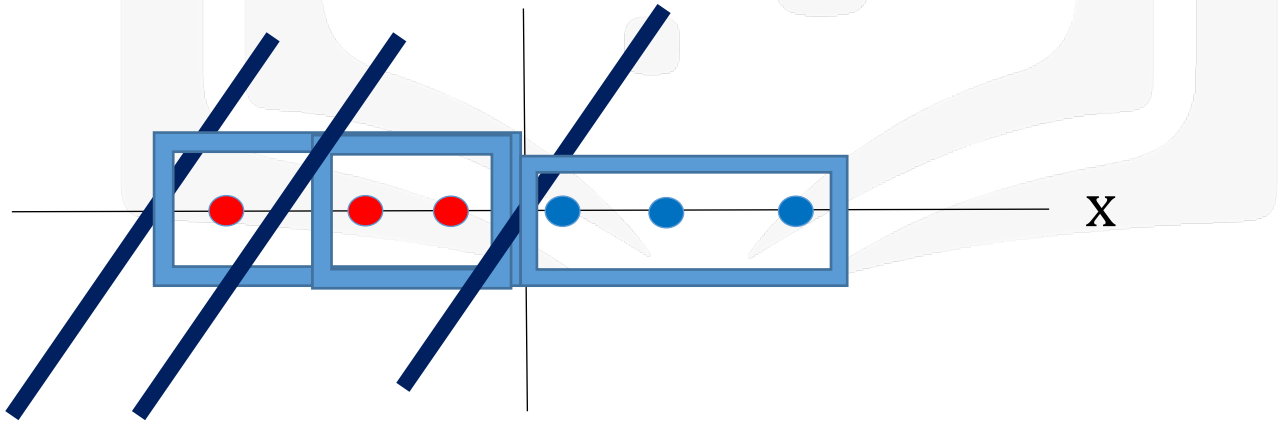
Training


$c(\boldsymbol{w}_0)$
$c(\boldsymbol{w}_1)$
$c(\boldsymbol{w}_2)$


$\boldsymbol{w}_0$
$\boldsymbol{w}_1$
$-\nabla f(\boldsymbol{\theta}_1)$
$\boldsymbol{w}_2$
$-\nabla f(\boldsymbol{\theta}_2)$
θ_2

Training



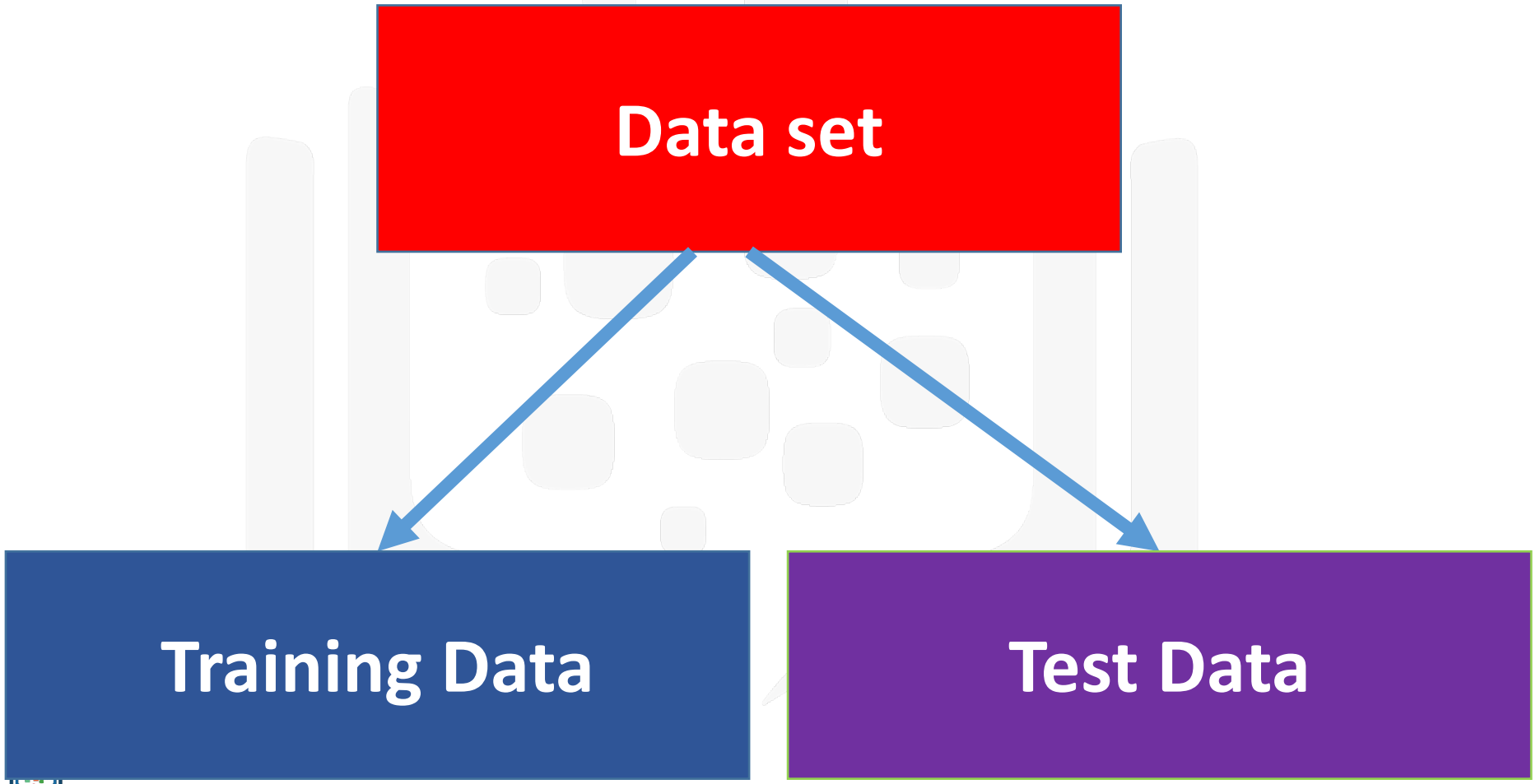


BIG DATA UNIVERSITY

© 2015 BigDataUniversity.com

Hold Out Validation

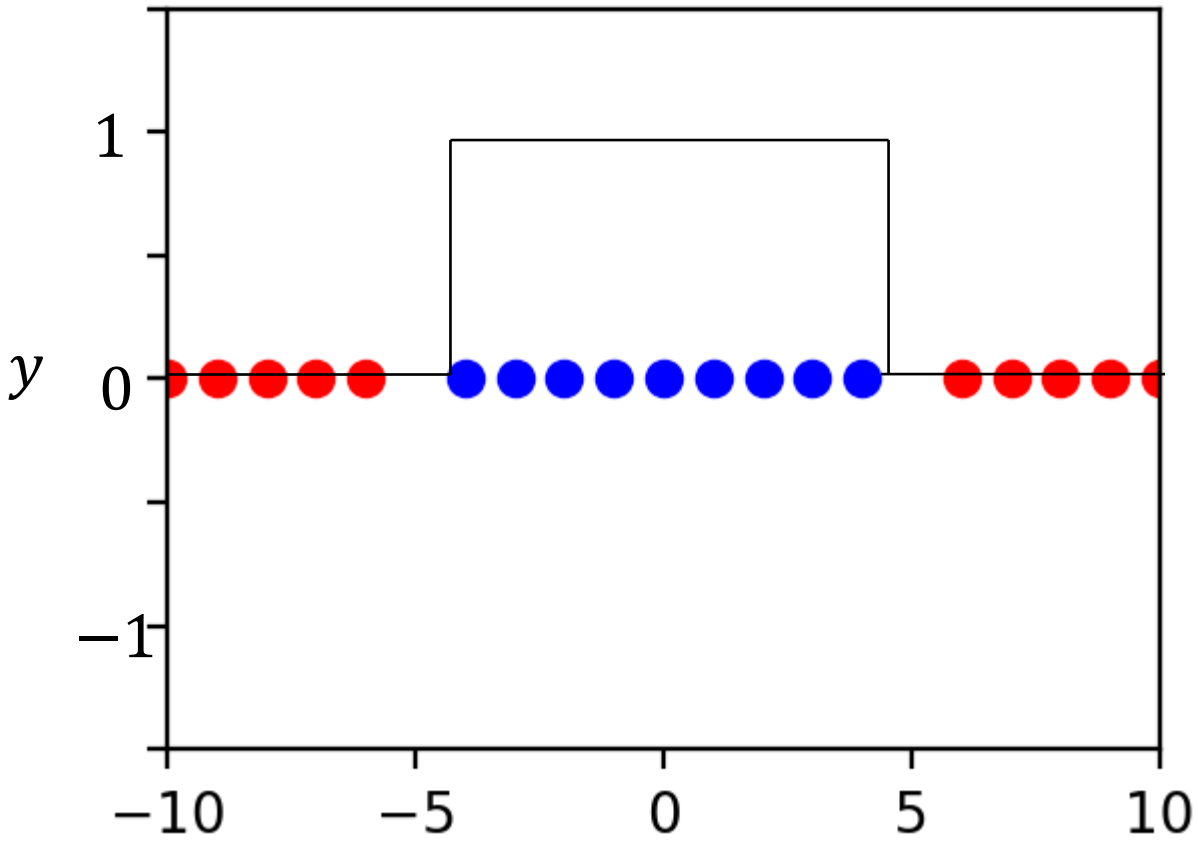
- Randomly Split the data into training and test set



© 2015 BigDataUniversity.com

Linear Classifiers vs Neural Networks

- This data set is not linearly Separable



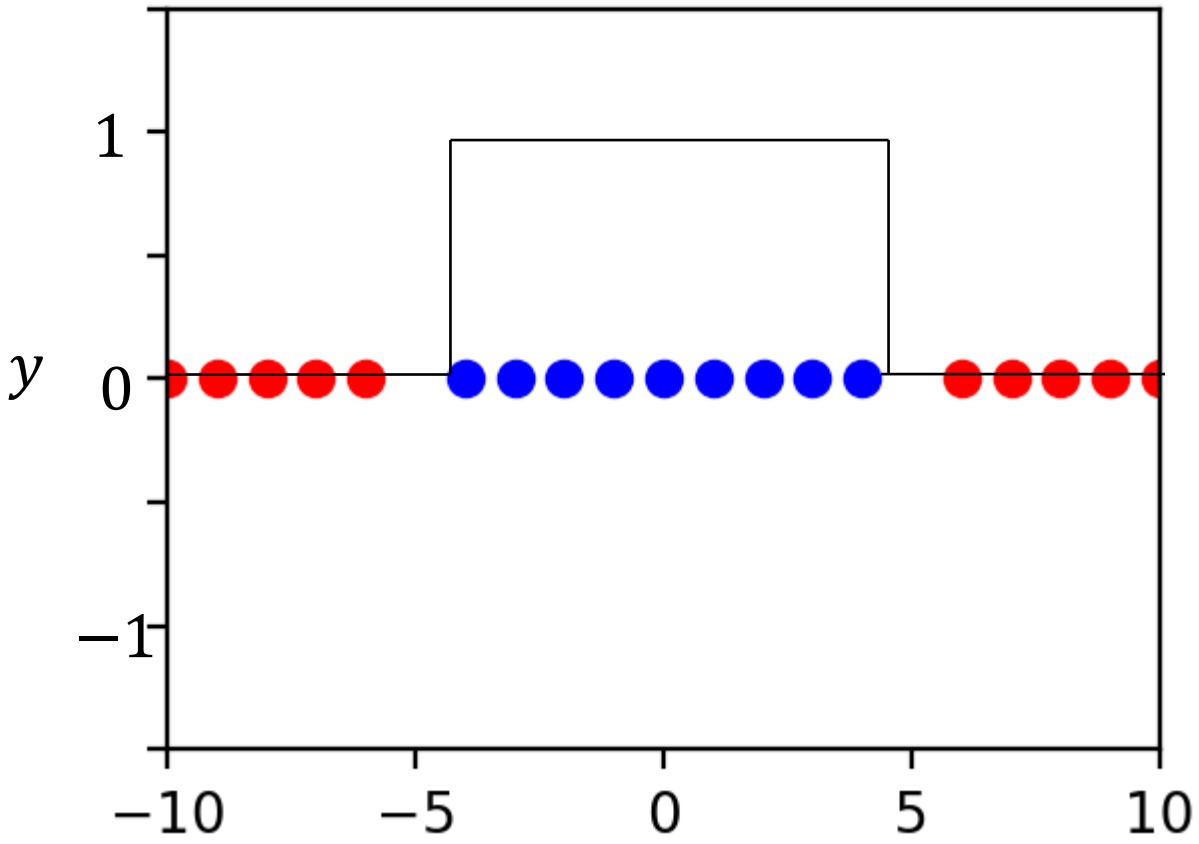
BIG DATA UNIVERSITY

Linear Classifiers vs Neural Networks

BIG DATA UNIVERSITY

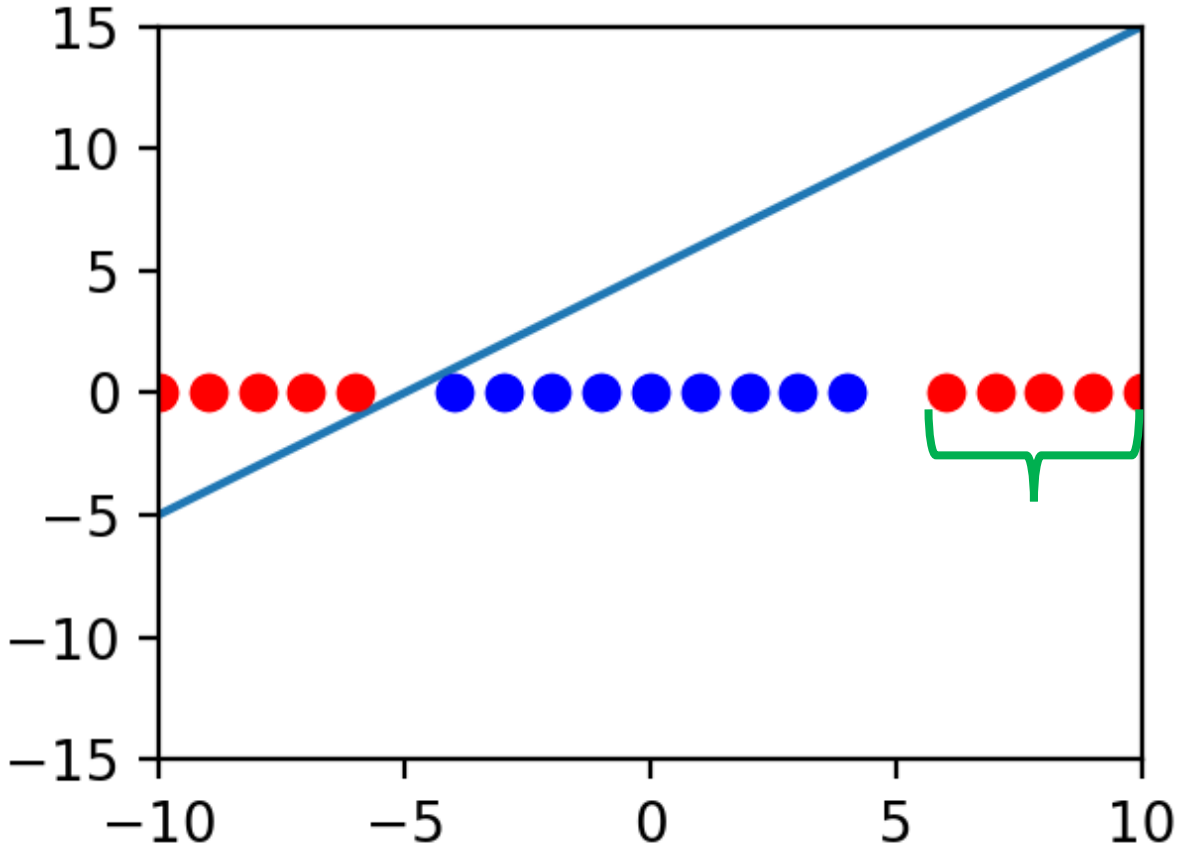
Linear Classifiers vs Neural Networks

- This data set is not linearly Separable



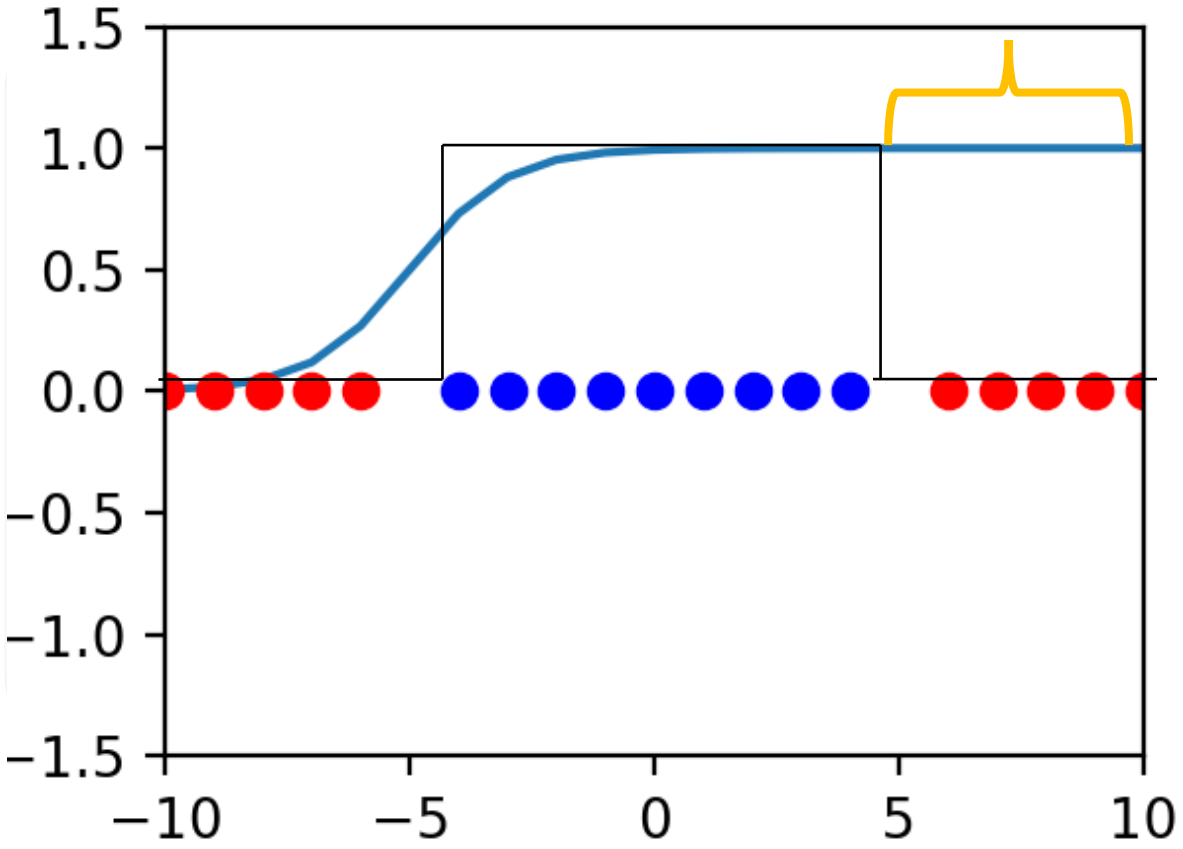
BIG DATA UNIVERSITY

Neural Networks



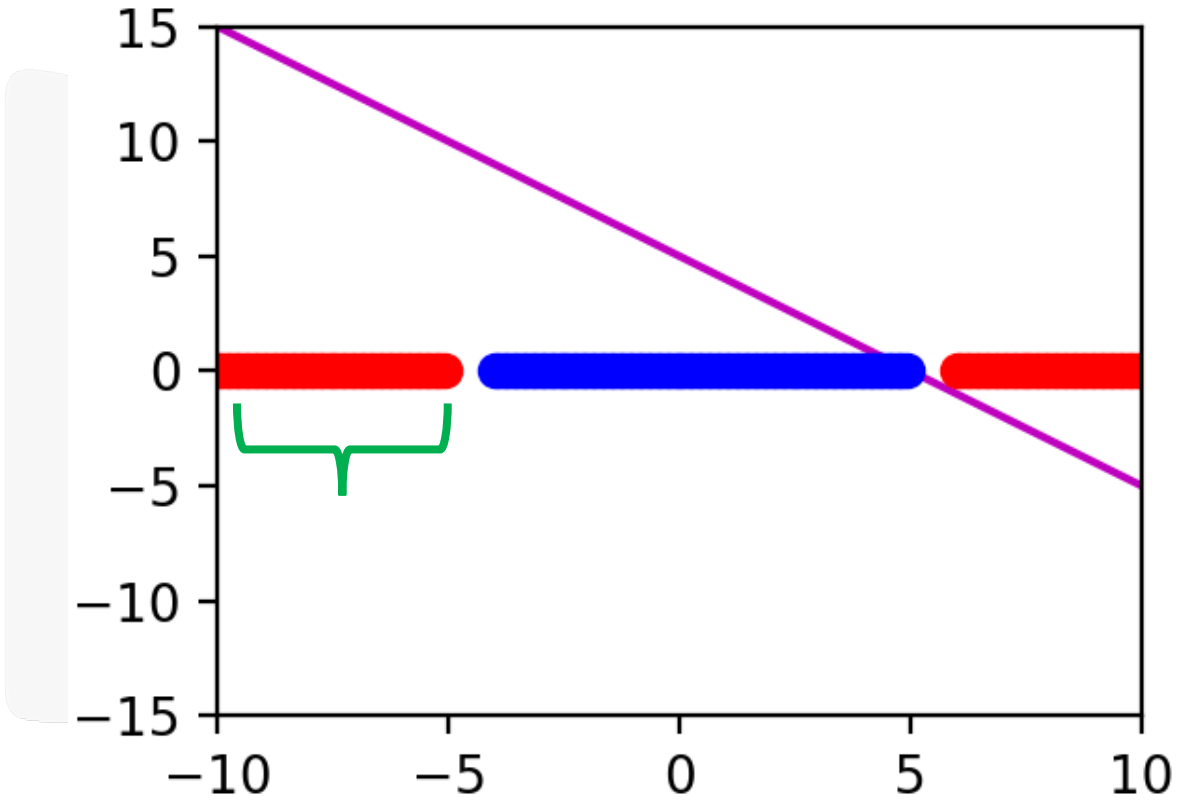
BIG DATA UNIVERSITY

Neural Networks



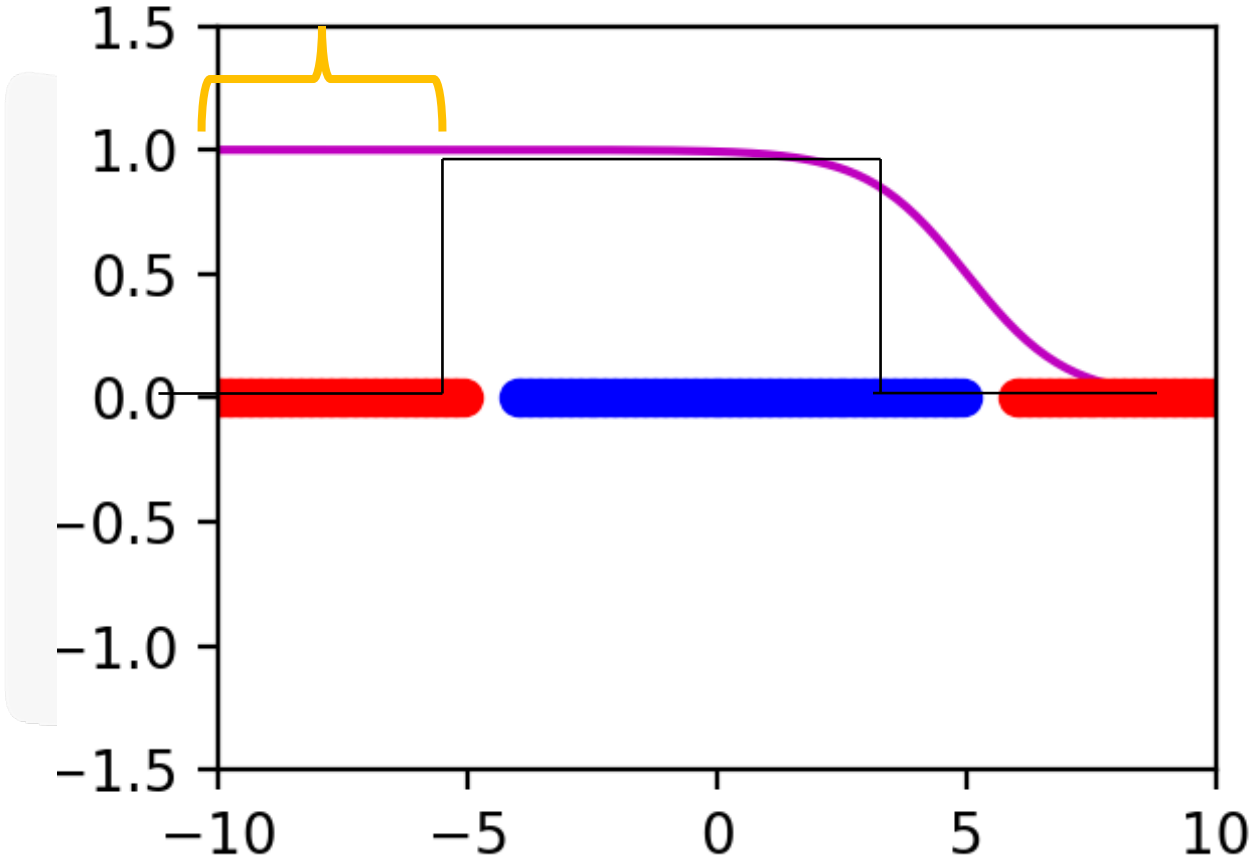
BIG DATA UNIVERSITY

Neural Networks



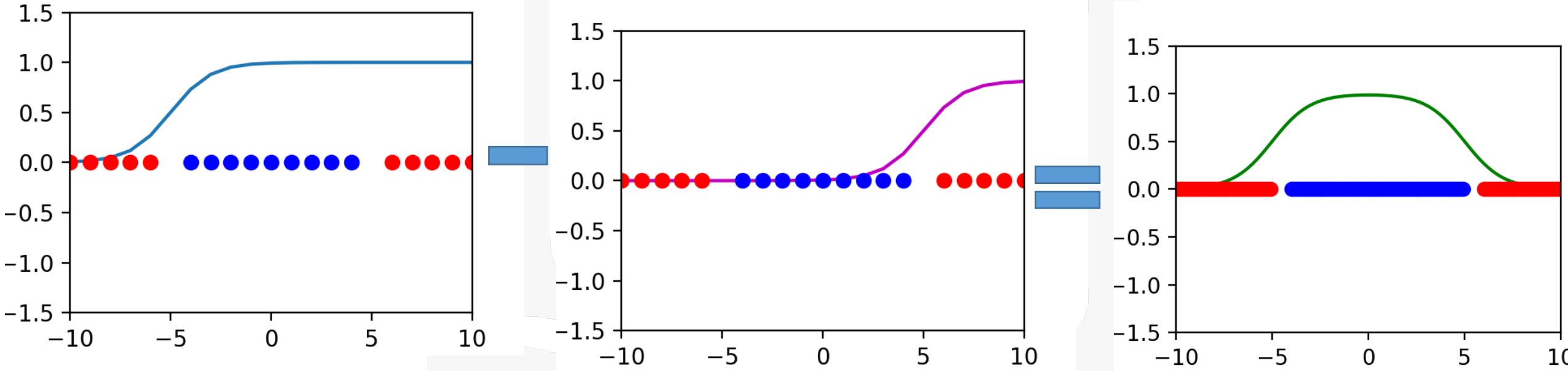
BIG DATA UNIVERSITY

Neural Networks



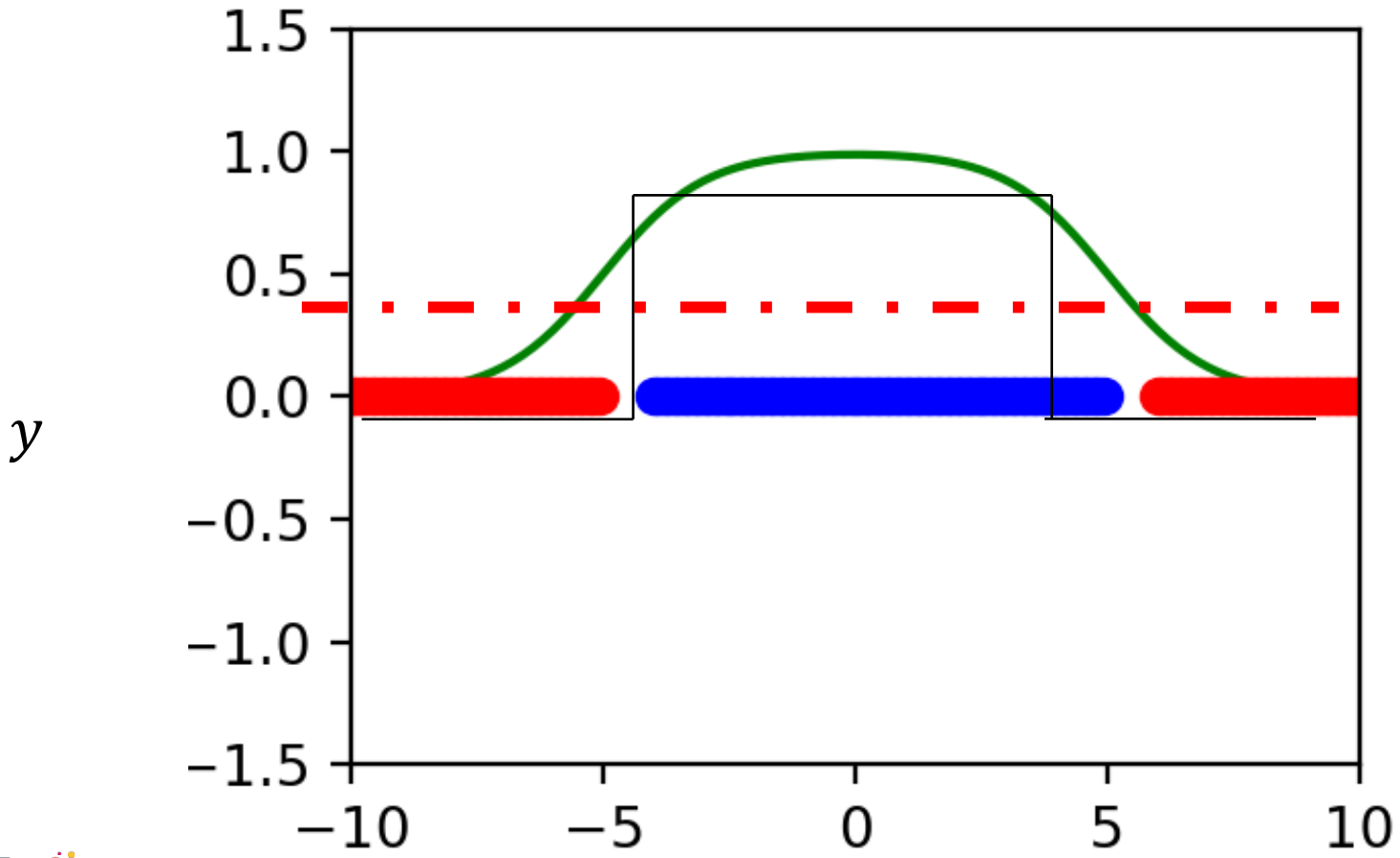
BIG DATA UNIVERSITY

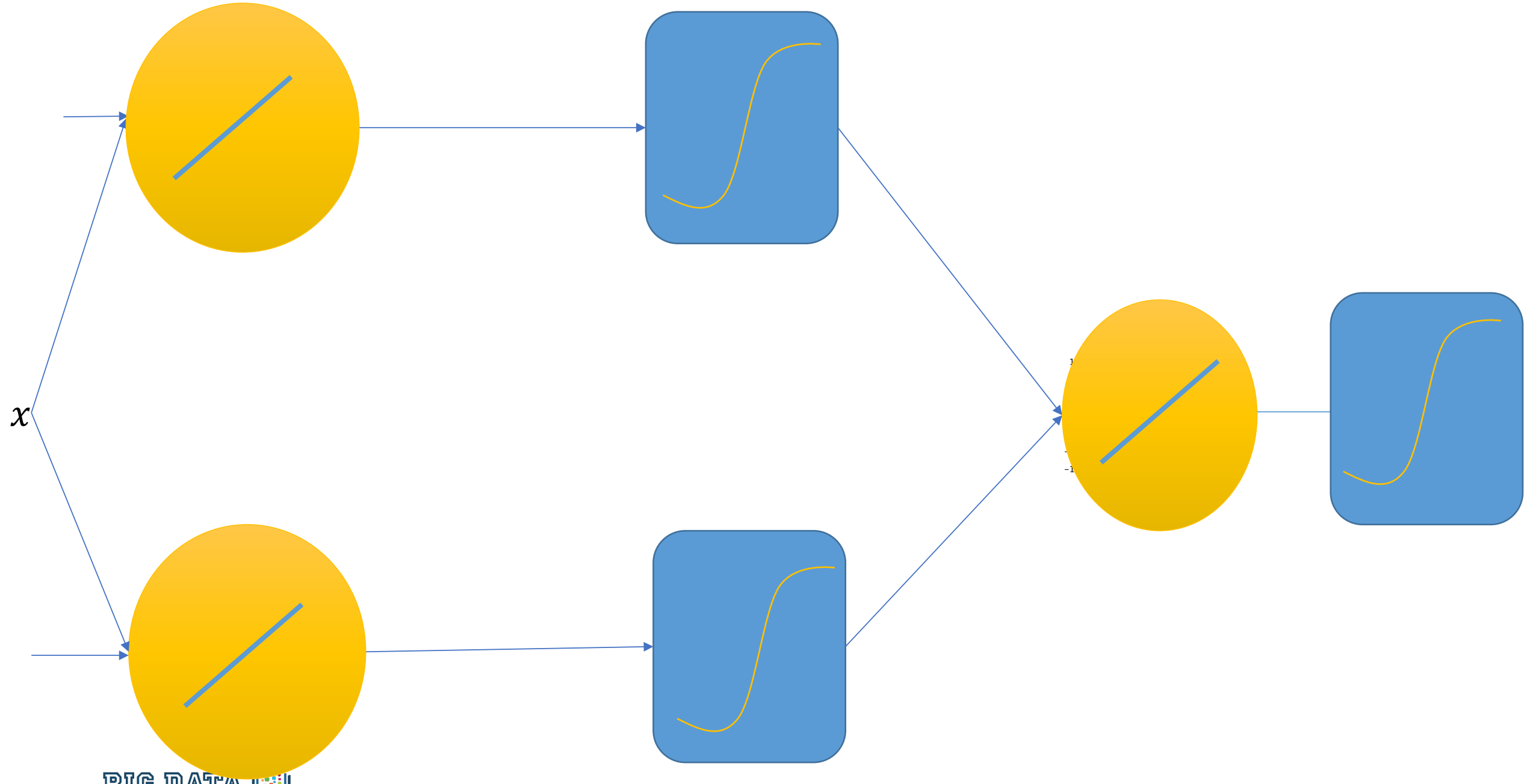
Neural Networks



BIG DATA UNIVERSITY

Neural Networks

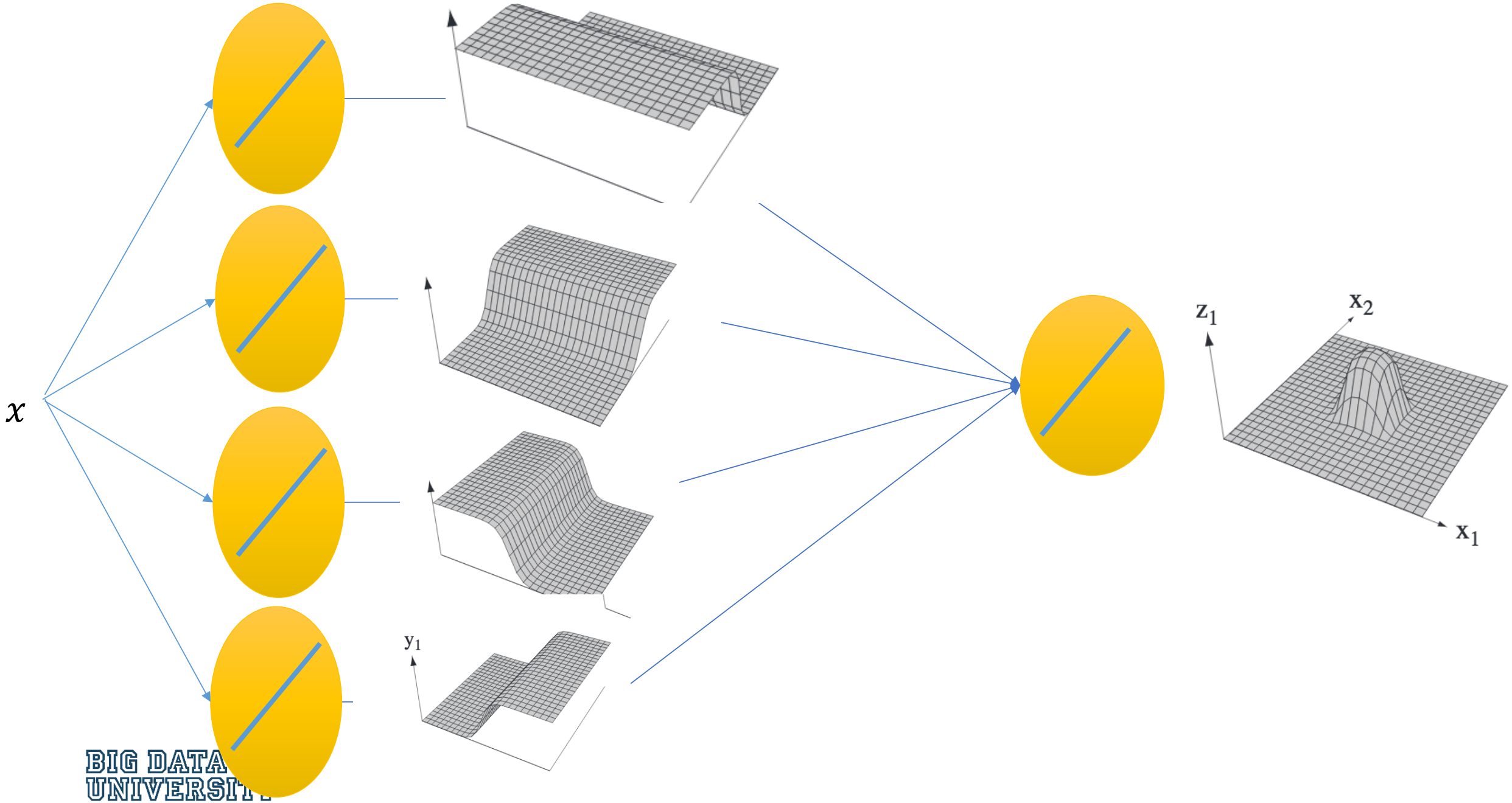



x
BIG DATA UNIVERSITY

Neural Network

- This is an example of a two layer network as you increase the with of the network you can construct arbitrary functions in any dimension
- Allowing us to make more complex decision function

BIG DATA UNIVERSITY

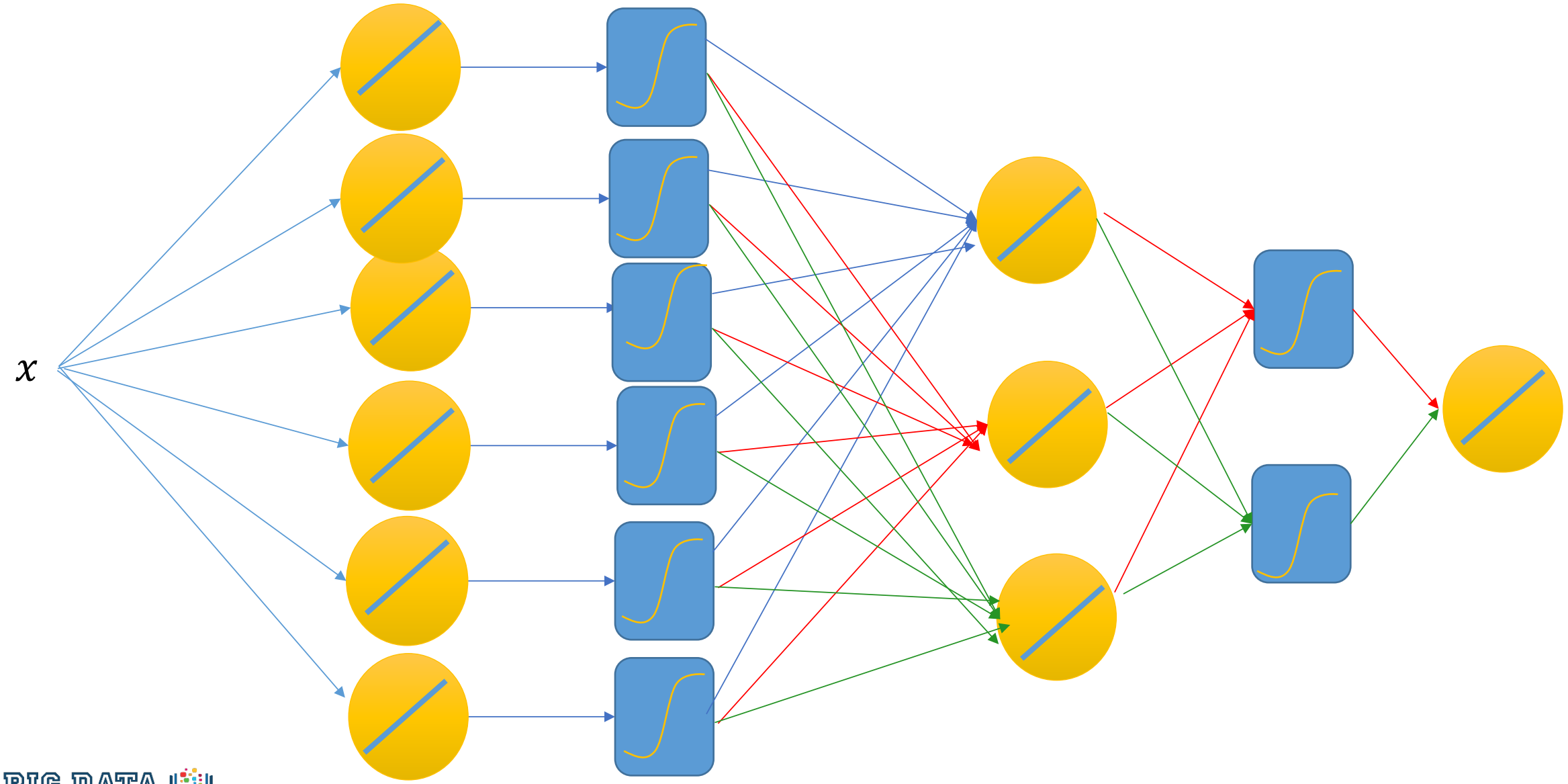


BIG DATA
UNIVERSITY

Linear Classifiers vs Neural Networks

- May get perfect accuracy on test data
- It turn out constructing arbitrary function is not learning
- We can also increase the number of layers i.e: going Deep

BIG DATA UNIVERSITY



BIG DATA UNIVERSITY