# CS 129.18

*Logistic* Regression

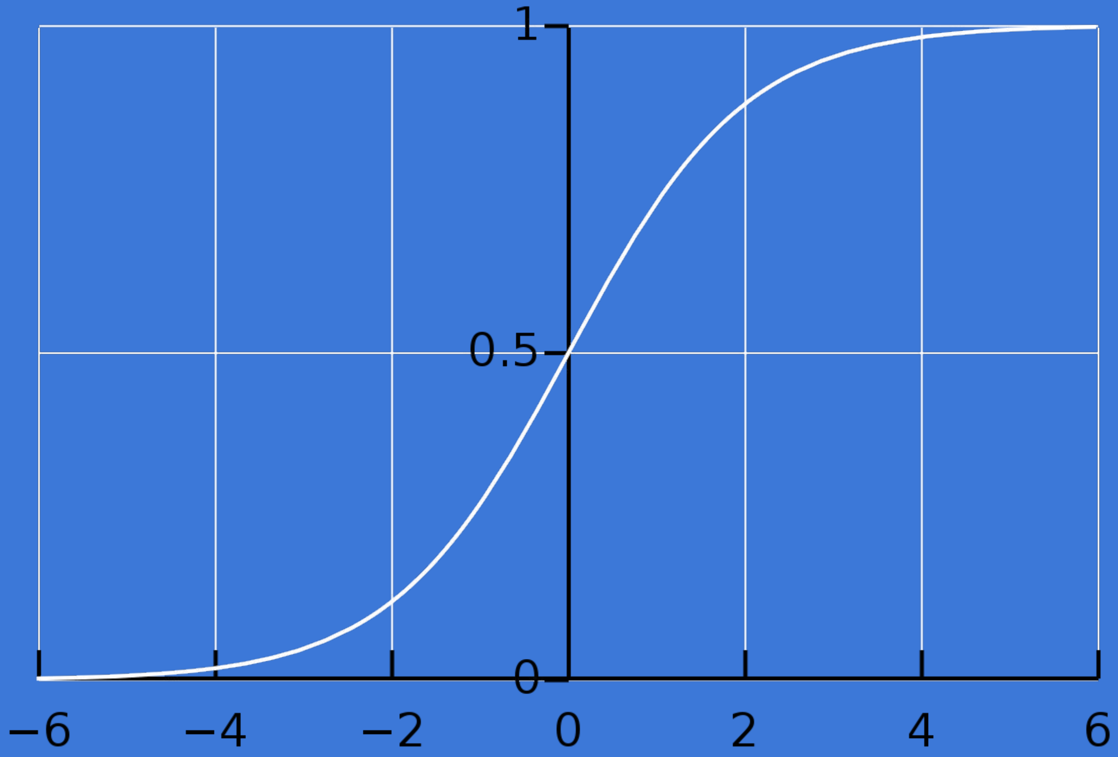
# Logistic Regression

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

# Sigmoid Function

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

The sigmoid function is perfect because it will output a value between 0 and 1




The function is asymptotic to 0 and 1, perfect for computational tasks

```python
def sigmoid(z):
    return 1/(1 + np.exp(-z))
```

Try it out for yourselves. It scales large numbers asymptotic to 1, and small ones to 0.

# Logistic Regression Function

$$h_{\Theta}(x) = g(\Theta^T x)$$

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

Pretty much the same as linear regression, but with the sigmoid function

$\Theta$ is the coefficient vector term for **logistic regression**

$\beta$ is the coefficient vector term for **linear regression**

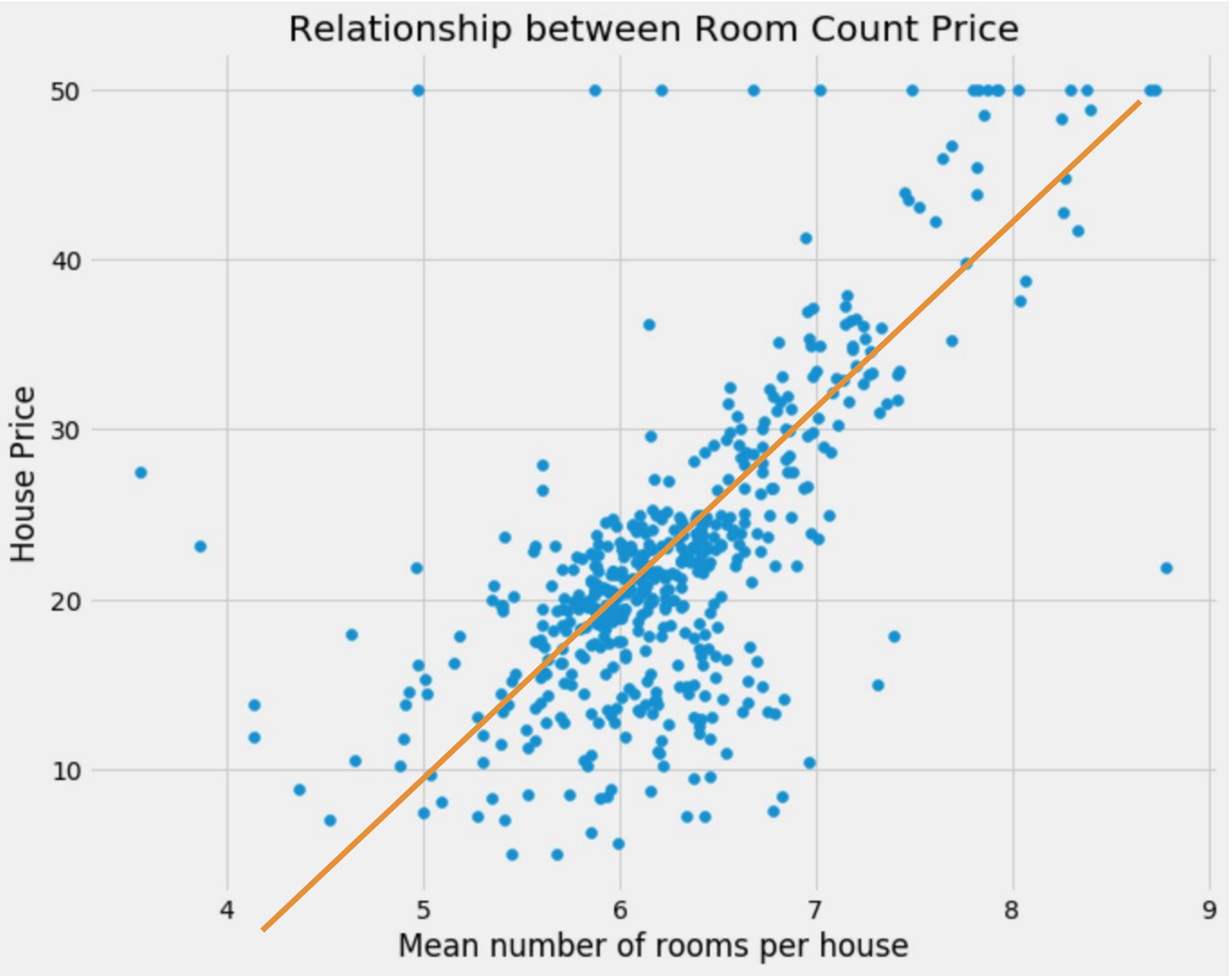
Because computer science and stat people disagree on notation

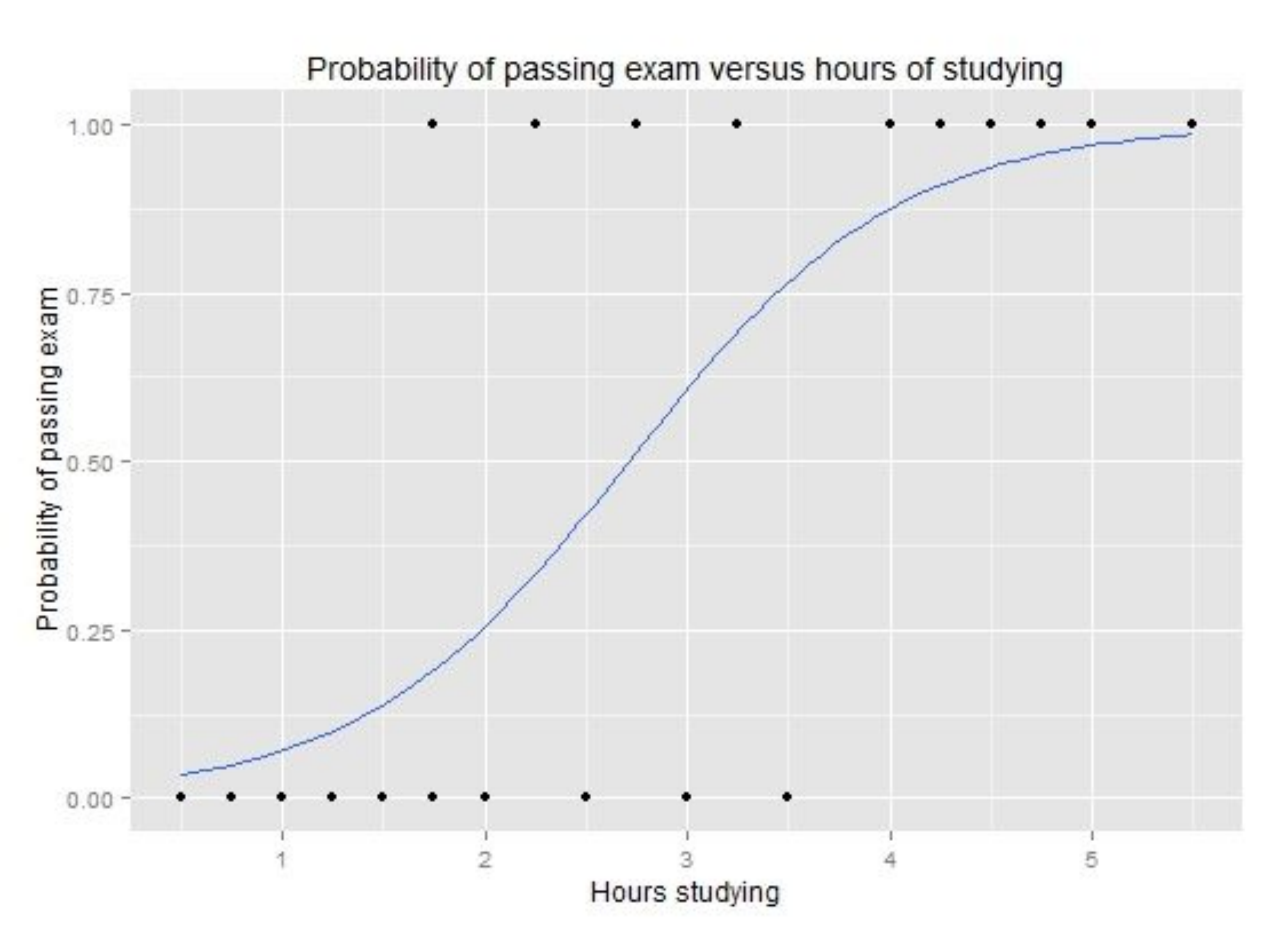
$$\Theta = \begin{bmatrix} \Theta \\ {}_0\Theta \\ 1\downarrow \\ \Theta \end{bmatrix}^T_{n}$$

$\Theta$ is the coefficient vector

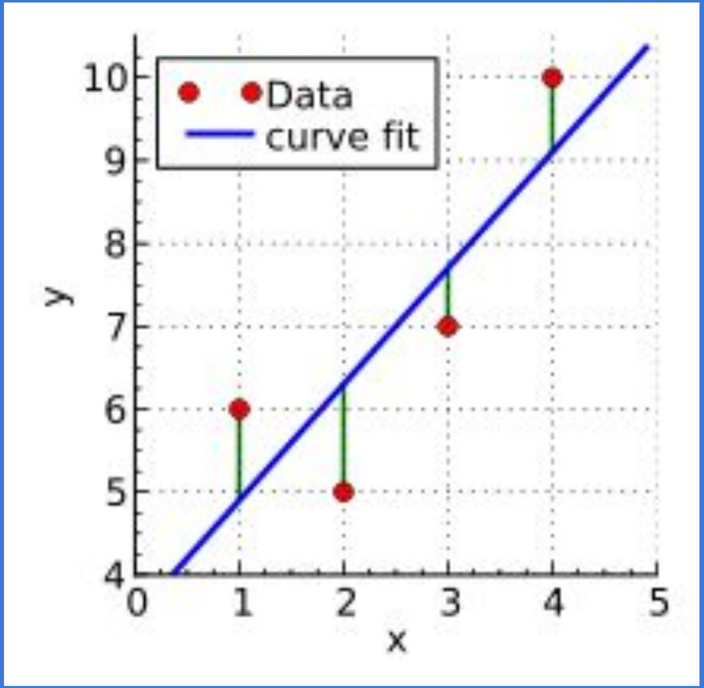
$$\mathbf{X} = \begin{bmatrix} x_0 \\ x_1 \\ \downarrow \\ x_n \end{bmatrix}^T$$

$\mathbf{X}$ is a vector of features you want to predict with

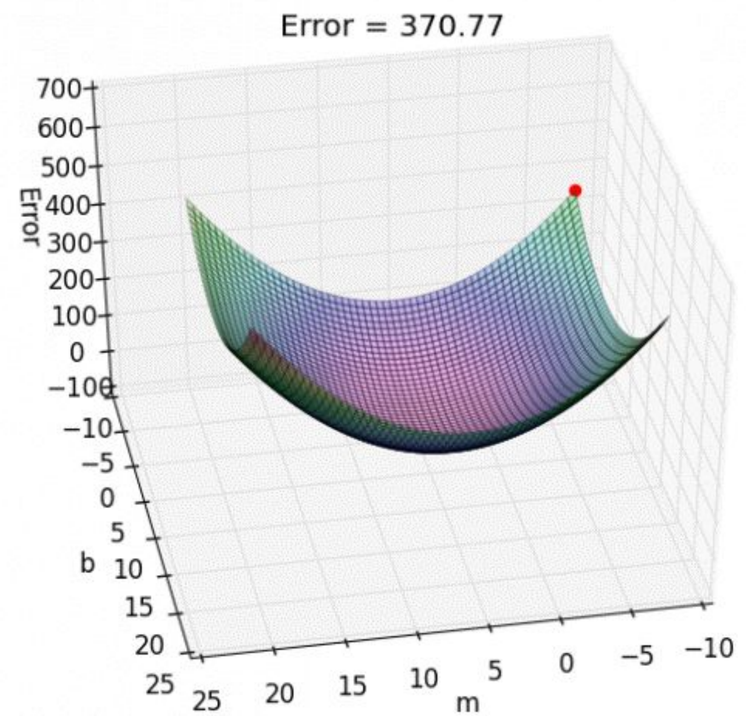

Relationship between Room Count Price
House Price
50
40
30
20
10
4
5
6
7
8
9
Mean number of rooms per house



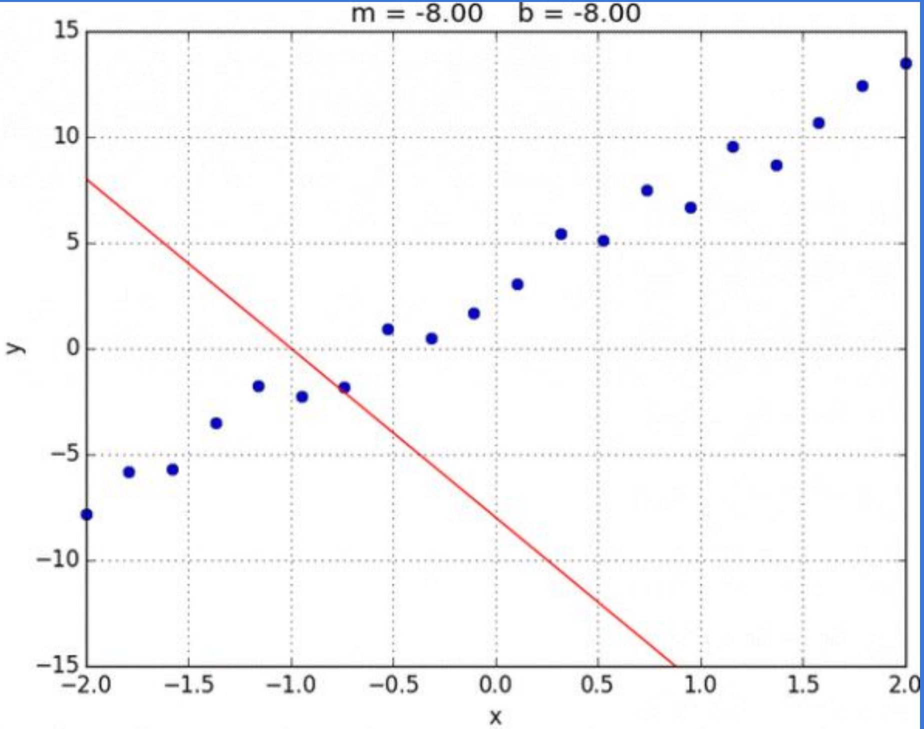
Probability of passing exam versus hours of studying
Probability of passing exam
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
Hours studying



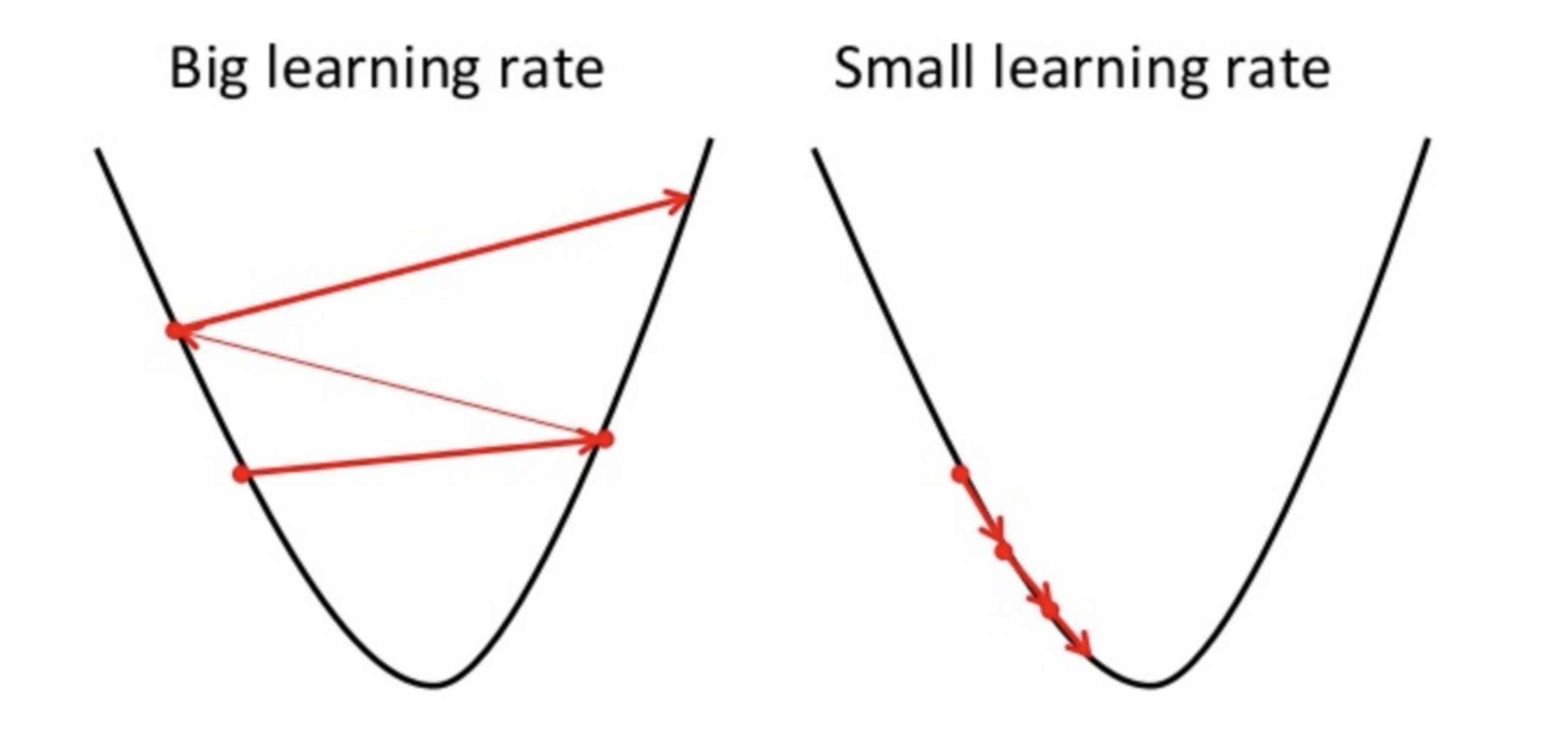

In Linear Regression, the green lines are your LOSS.



Error = 370.77
Error
700
600
500
400
300
200
100
0
−100
b
−10
−5
0
5
10
15
20
25
m
25
20
15
10
5
0
−5
−10

m = -8.00 b = -8.00
y
15
10
5
0
−5
−10
−15
x
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0


Big learning rate
Small learning rate

# Gradient Descent

$$
\begin{aligned}
&\textit{Repeat } \{ \\
&\quad \theta_j := \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \\
&\}
\end{aligned}
$$

# You get to classify binary datasets

# Thank you