Example - A customer purchase decision for onion during her visits to a vegetable store are shown below in the table.

Visit Index	1	2	3	4	5	6	7	8	9	10
Price	2	2	2	2	2	3	3	3	3	4
Decision	Y	Y	Y	N	Y	Y	Y	N	N	Y

Visit Index	11	12	13	14	15	16	17	18	19	20
Price	4	4	4	4	4	5	5	5	5	5
Decision	N	N	Y	N	N	N	N	N	N	Y

Logistic Regression

$$\left(p_2\ p_2\ p_2(1-p_2)p_2\right)\left(p_3^2(1-p_3)^2\right)\left(p_4^2(1-p_4)^4\right)$$

Example - A customer purchase decision for onion during her visits to a vegetable store are shown below in the table.

$$p_5(1-p_5)^4$$

Visit Index	1	2	3	4	5	6	7	8	9	10
Price	2	2	2	2	2	3	3	3	3	4
Decision	Y	Y	Y	N	Y	Y	Y	N	N	Y

Visit Index	11	12	13	14	15	16	17	18	19	20
Price	4	4	4	4	4	5	5	5	5	5
Decision	N	N	Y	N	N	N	N	N	N	Y

When the price is 2.5, what is the probability that the person will make a purchase?

When the price is 2.5, what is the probability that the person will make a purchase?

Price	No. of Visits	No. of Purchases	Probability of Purchase
2	5	4	0.8
3	4	2	0.5
4	6	2	0.33
5	5	1	0.2

Example Contd-

When the price is 6, what is the probability that the person will make a purchase?

Price	No. of Visits	No. of Purchases	Probability of Purchase
2	5	4	0.8
3	4	2	0.5
4	6	2	0.33
5	5	1	0.2

Ideas to answer the problem

Objective - Find the probability of purchase as a function of price

$P(\text{purchase}) = f(\text{Price})$

Can we make use of linear regression??

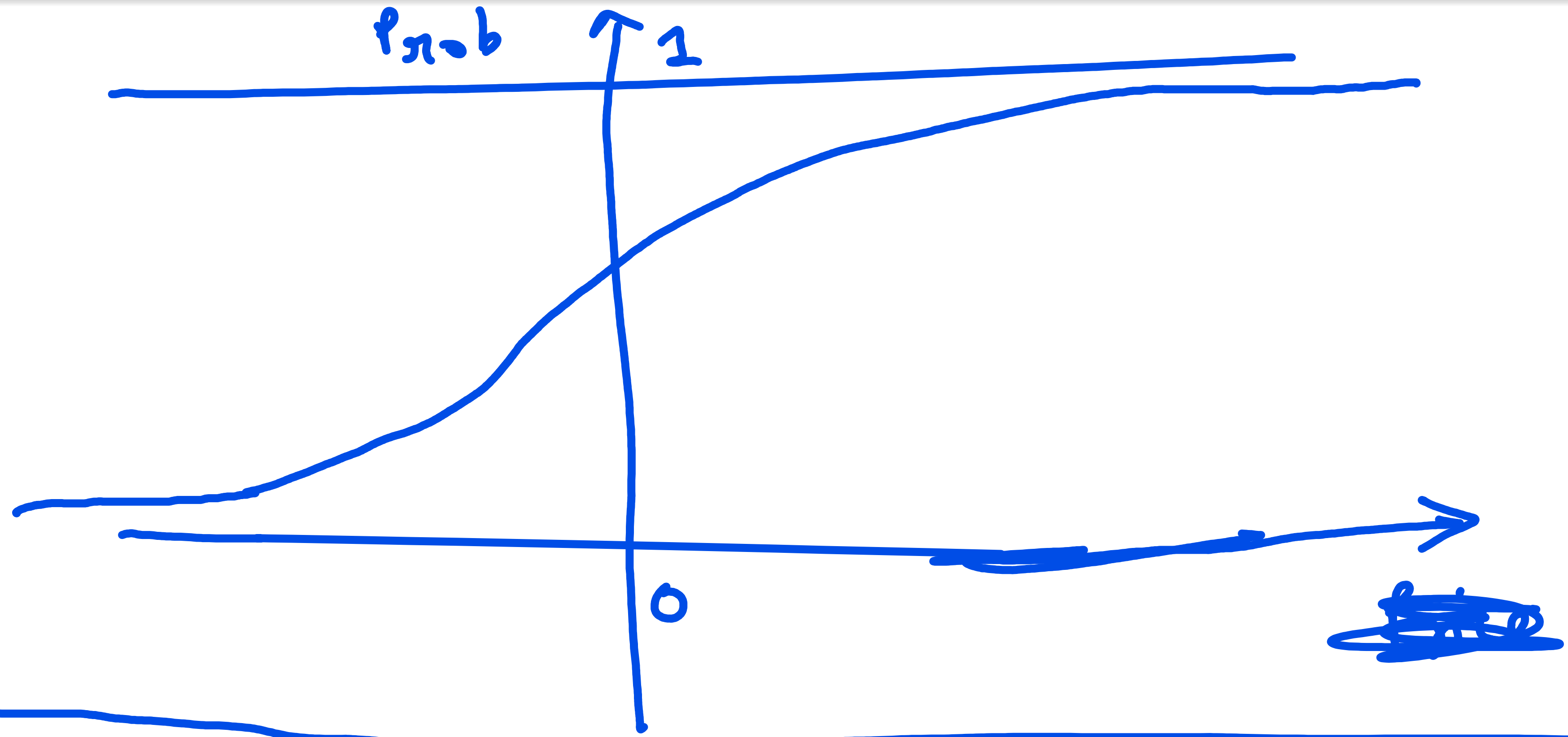
Logistic regression

Attempt for a non-linear fit???

Limitations of Linear Regression

- Doesn't explicitly recognize 0-1 nature of the response
- The impact of change in price on probability of purchase decision is different at different levels of price
- Assumed Linear regression equation treats it as a constant

Logistic Regression - Main Ideas



$$\frac{e^x}{1+e^x} = y$$

$$\frac{dy}{dx} = y(1-y)$$

$$\frac{e^{-x}}{(1+e^{-x})}\left[\frac{1}{(1+e^{-x})}\right]$$

Maximum Likelihood Principle - Ideas

Rain Prediction Model				
	Monday	Tuesday	Wednesday	Thursday
Model A	0.3	0.6	0.7	0.2
Model B	0.2	0.4	0.9	0.6
Rained?	N	Y	Y	N

Model C 0.1 0.6 0.6 0.3

Which Model is better??

Model D 0.1 0.9 0.9 0.1

Maximum Likelihood Principle - Example

Toss a coin 100 times

You get 70 Heads.

[H H T H H T H T T T T H H H _ _ _ _ _ _ T H]

What "prob" of Heads will maximize the chances of observing what you observed ??

$$\boxed{p_H^{70}(1-p_H)^{30}}$$

Maxi:

$$70\, p_H^{69}(1-p_H)^{30} + (-30)\, p_H^{70}(1-p_H)^{29} = 0$$

Logistic Regression - Implementation in Python

Logistic Regression - Multiple Classes

Multinomial Regression

Multinomial Regression - Softmax Function

Softmax Function - Implementation in Python