به نام خدا

کارگاه علم داده با پایتون پیشرفته

جلسه پنجم: رگرسیون لجستیک

مدرس :

مهرناز جلیلی

دانشجو کارشناسی ارشد علم داده ها

دانشگاه شهید بهشتی

Classification

Logistic Regression/ Intro

- who is leaving and why
- Close to Regression but here, Y is a categorical (here binary) value
- All Xs should be continues, or converted to “continues”

	tenure	age	address	income	ed	employ	equip	callcard	wireless	churn
0	11.0	33.0	7.0	136.0	5.0	5.0	0.0	1.0	1.0	1
1	33.0	33.0	12.0	33.0	2.0	0.0	0.0	0.0	0.0	1
2	23.0	30.0	9.0	30.0	1.0	2.0	0.0	0.0	0.0	0
3	38.0	35.0	5.0	76.0	2.0	10.0	1.0	1.0	1.0	0
4	7.0	35.0	14.0	80.0	2.0	15.0	0.0	1.0	0.0	0

Classification

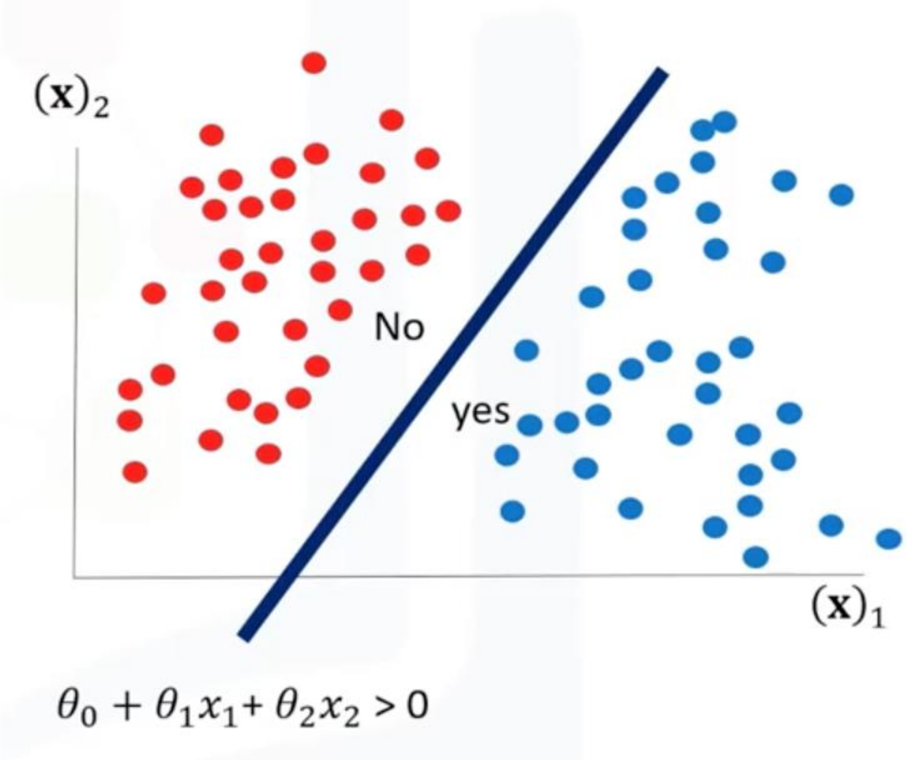
Logistic Regression/ Intro

- Predicting a disease
- chance of mortality based on a situation
- halting a subscription
- purchase
- failure of a product
- ...

Classification

Logistic Regression/ Intro

- Target should be category (or better, binary)
- We need the probability of prediction
- we need a linear decision boundary (line or even polynomial)
- We need to understand the impact of features (Theta is closer to 0 or is high)



Classification

Logistic Regression vs Linear Regression

$$\overbrace{\text{tenure} \quad \text{age} \quad \text{address} \quad \text{income} \quad \text{ed} \quad \text{employ} \quad \text{equip} \quad \text{callcard} \quad \text{wireless}}^{\mathbf{X}} \quad \overbrace{\text{churn}}^{y}$$

	tenure	age	address	income	ed	employ	equip	callcard	wireless	churn
0	11.0	33.0	7.0	136.0	5.0	5.0	0.0	1.0	1.0	1.0
1	33.0	33.0	12.0	33.0	2.0	0.0	0.0	0.0	0.0	1.0
2	23.0	30.0	9.0	30.0	1.0	2.0	0.0	0.0	0.0	0.0
3	38.0	35.0	5.0	76.0	2.0	10.0	1.0	1.0	1.0	0.0

$$X \in \mathbb{R}^{m \times n}$$

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

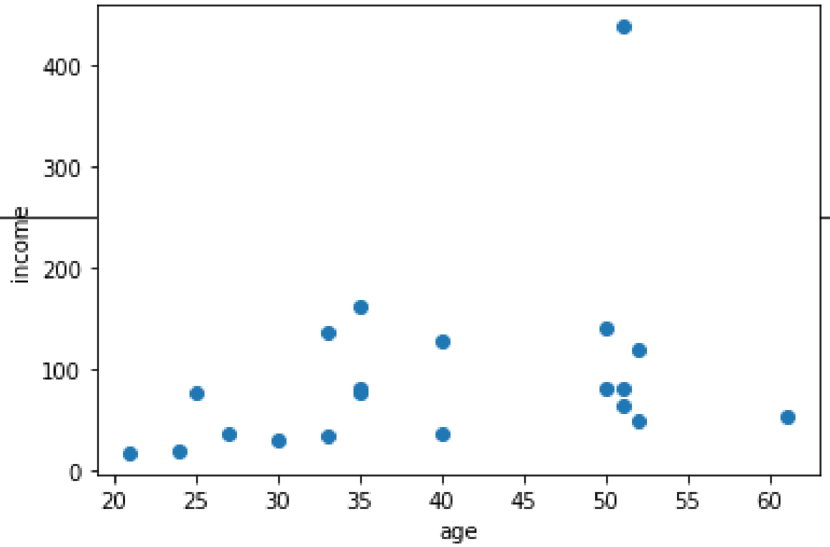
$$\boxed{\hat{y} = \mathrm{P(y=1|x)}}$$

$$\mathrm{P(y=0|x) = 1- P(y=1|x)}$$

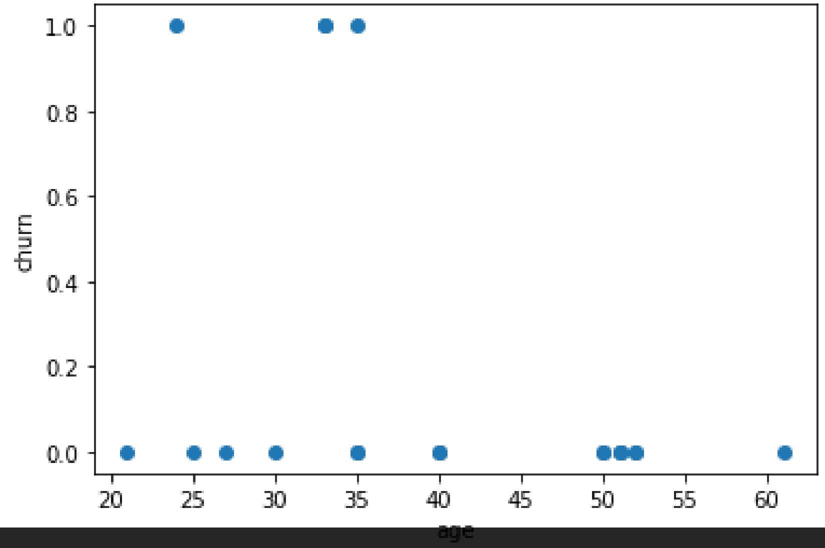
Classification

Logistic Regression vs Linear Regression

- On Previous data, try linear with age vs income



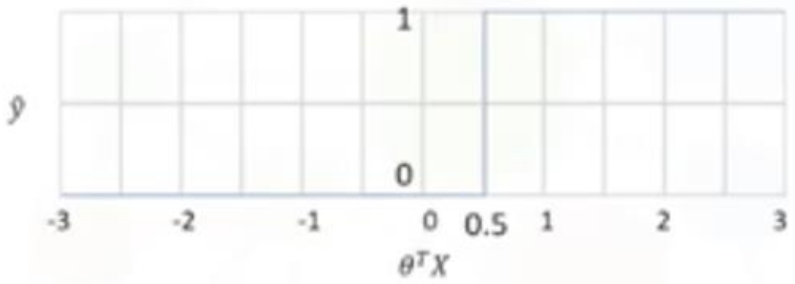
- now repeat, trying with age vs churn: funny and we should have a step function as threshold



Classification

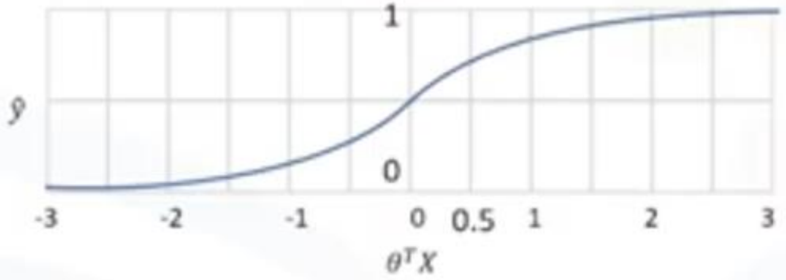
Logistic Regression vs Linear Regression / Sigmoid

$$\theta^T X = \theta_0 + \theta_1 x_1 + \cdots$$



$$\hat{y} = \begin{cases} 0 & \text{if } \theta^T X<0.5 \\ 1 & \text{if } \theta^T X \geq 0.5 \end{cases}$$

$$\sigma(\theta^T X) = \sigma(\theta_0 + \theta_1 x_1 + \cdots)$$



$$\hat{y} = \sigma(\theta^T X)$$

P(y=1|x)

Classification

Logistic Regression Training

$\sigma(\theta^T X) \longrightarrow$ P(y=1|x)

1. Initialize θ.
2. Calculate $\hat{y} = \sigma(\theta^T X)$ for a customer.
3. Compare the output of $\hat{y}$ with actual output of customer, y, and record it as error.
4. Calculate the error for all customers.
5. Change the θ to reduce the cost.
6. Go back to step 2.

$\theta = [-1,2]$

$\hat{y} = \sigma([-1,2] \times [2,5]) = 0.7$

Error = 1-0.7 = 0.3

$Cost = J(\theta)$

θ_{new}

Classification

Logistic Regression Training

- Cost Function
- we have to minimize the Cost
- Can be done via derivative but its difficult

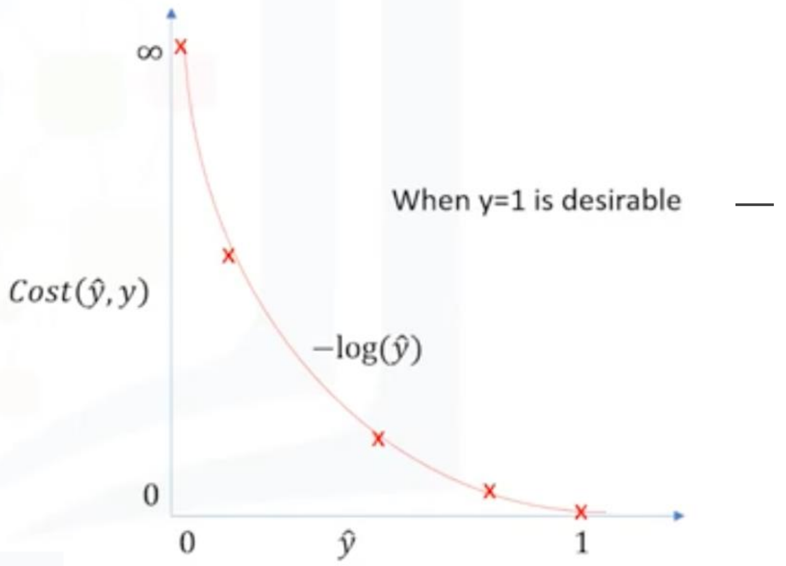
$$Cost(\hat{y}, y) = \frac{1}{2}(\sigma(\theta^T X) - y)^2$$

$$J(\theta) = \frac{1}{m}\sum_{i-1}^{m} Cost(\hat{y}, y)$$

Classification

Logistic Regression Training

- We can define a new Cost function!
- here there are more approaches to minimize the function; say Gradient Descent (iterative technique)



$$Cost(\hat{y}, y) = \frac{1}{2}\left(\sigma(\theta^T X) - y\right)^2$$

$$Cost(\hat{y}, y) = \begin{cases} -\log(\hat{y}) & if\ y = 1 \\ -\log(1-\hat{y}) & if\ y = 0 \end{cases}$$

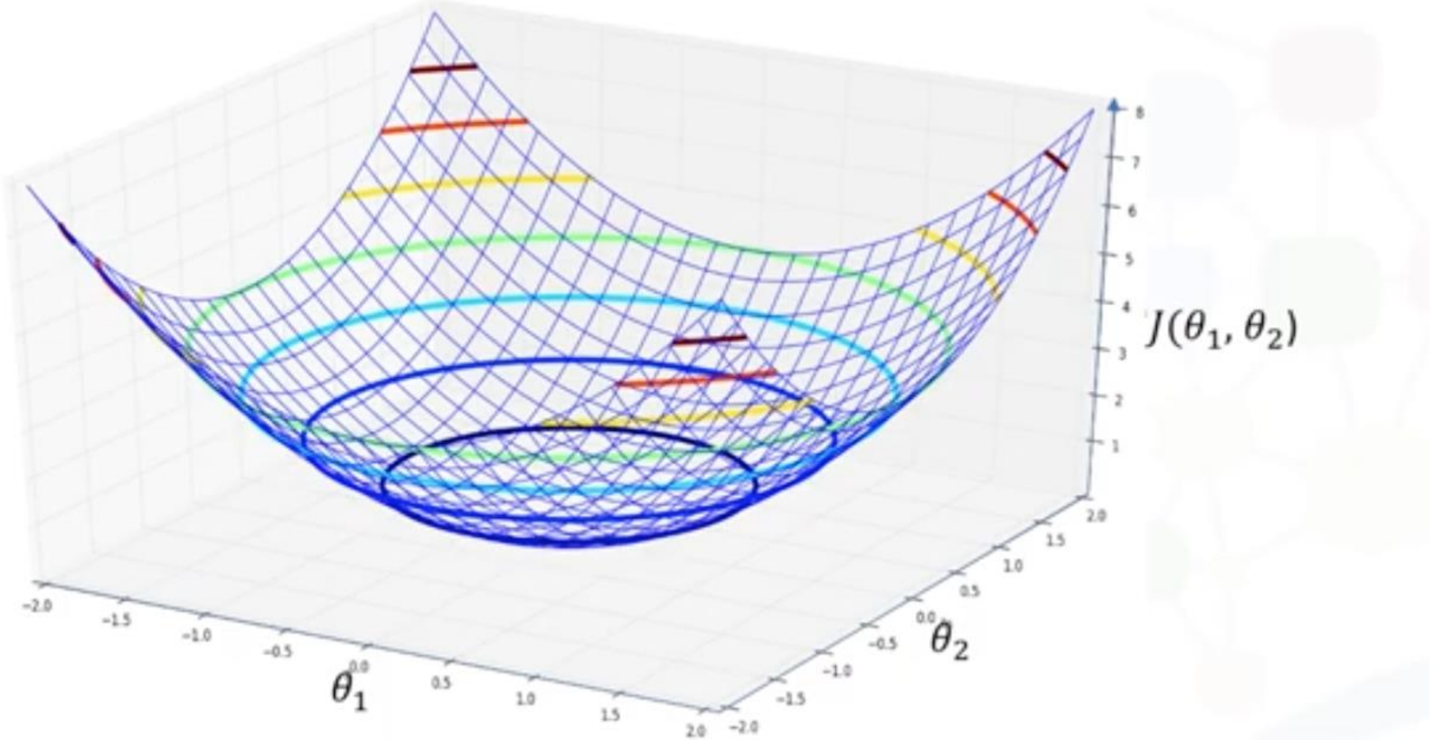
$$J(\theta) = \frac{1}{m}\sum_{i-1}^{m} Cost(\hat{y}, y)$$

$$J(\theta) = -\frac{1}{m}\sum_{i-1}^{m} y^i \log(\hat{y}^i) + (1-y^i)\log(1-\hat{y}^i)$$

Classification

Logistic Regression Training

- Gradient descent is an iterative approach to finding the minimum of a function. It uses the derivative of a cost function to change the parameter values to minimize the cost or error.



$$\hat{y} = \sigma(\theta_1 x_1 + \theta_2 x_2)$$

$$J(\theta) = -\frac{1}{m}\sum_{i-1}^{m} y^i \log(\hat{y}^i) + (1 - y^i)\log(1 - \hat{y}^i)$$

Classification

Logistic Regression Training

1. initialize the parameters randomly.
2. Feed the cost function with training set, and calculate the error.
3. Calculate the gradient of cost function.
4. Update weights with new values.
5. Go to step 2 until cost is small enough.
6. Predict the new customer X.

$$\theta^T = [\theta_0, \theta_1, \theta_2, ...]$$

$$J(\theta) = -\frac{1}{m}\sum_{i-1}^{m} y^i \log(\hat{y}^i) + (1 - y^i)\log(1 - \hat{y}^i)$$

$$\nabla J = [\frac{\partial J}{\partial \theta_1}, \frac{\partial J}{\partial \theta_2}, \frac{\partial J}{\partial \theta_3}, \ldots, \frac{\partial J}{\partial \theta_k}]$$

$$\theta_{new} = \theta_{prv} - \eta \nabla J$$

$$P(y=1|x) = \sigma(\theta^T X)$$

Lab: Logistic Regression

- L05-MehrnazJalili