Machine Learning Problems

Dr. Pritam Anand.

Assistant Professor,

DA-IICT, Gandhinagar.

Binary Classification Problem

Length (cm)	Height (cm)	Number of fins	Weight (Kg)	Color	Fish type
17.8	22.9	8	5.1`	1	1
14.8	20.5	7	4.9	2	-1
16. 34	12.76	6	6.6	3	1
10. 34	8.76	3	3.8	3	1
---	-----	----	------	------	--------
11 .30	17.76	6	9.8	1	-1

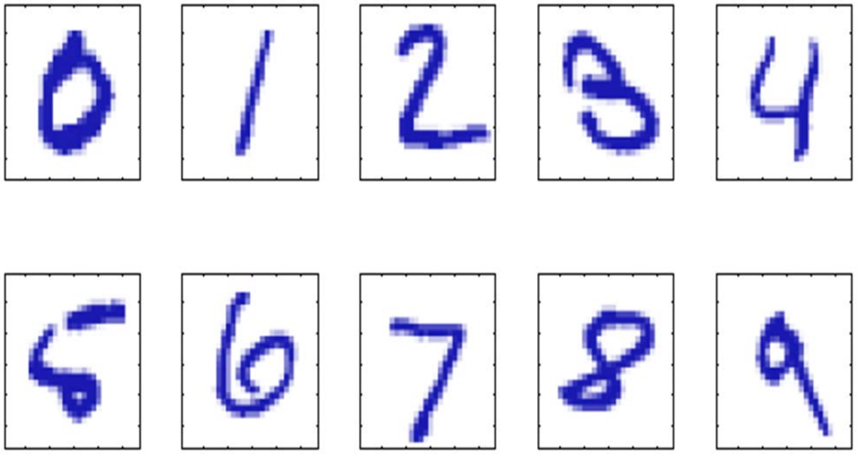
Binary Classification :-

Given a training set $\{ (X_i , Y_i) : X_i \in R^n , Y_i \in \{-1,1\} , i = 1,2,.., l \}$, find a function f which can efficiently predict the label Y for an unseen $X \in R^n$.

Multi class problems

Multi-class Classification :-

Given a training set $\{ (X_i , Y_i) : X_i \in R^n , Y_i \in \{1,2,...k\} , i = 1,2,.., l \}$, find a function f which can efficiently predict the label Y for an unseen $X \in R^n$.

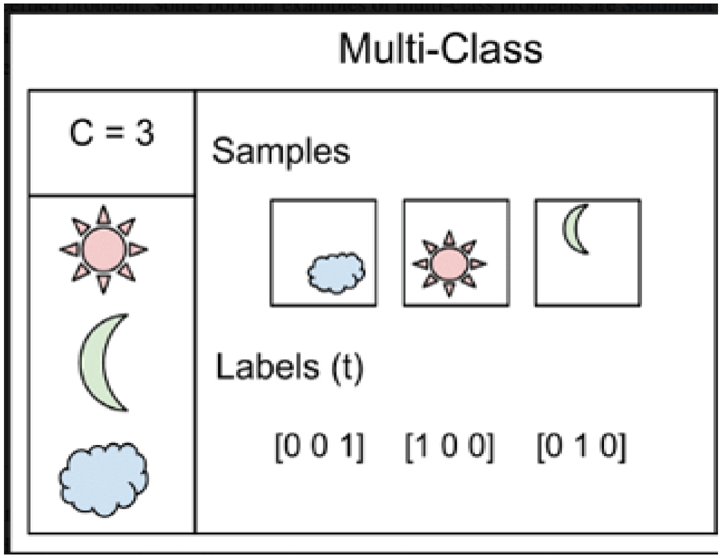


Multi class problems

Multi-class Classification :-

Given a training set $\{ (X_i , Y_i) : X_i \in R^n , Y_i \in \{1,2,\ldots k\} , i = 1,2,.., l \}$, find a function f which can efficiently predict the label Y for an unseen $X \in R^n$.

Image Content Classification

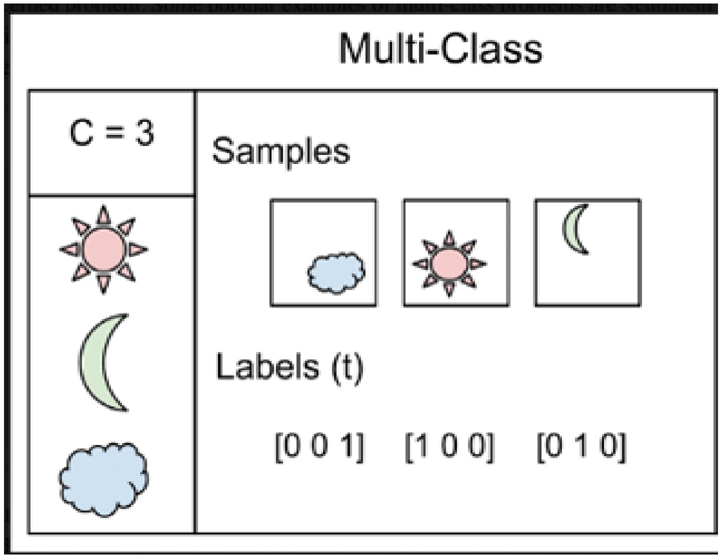


Multi class problems

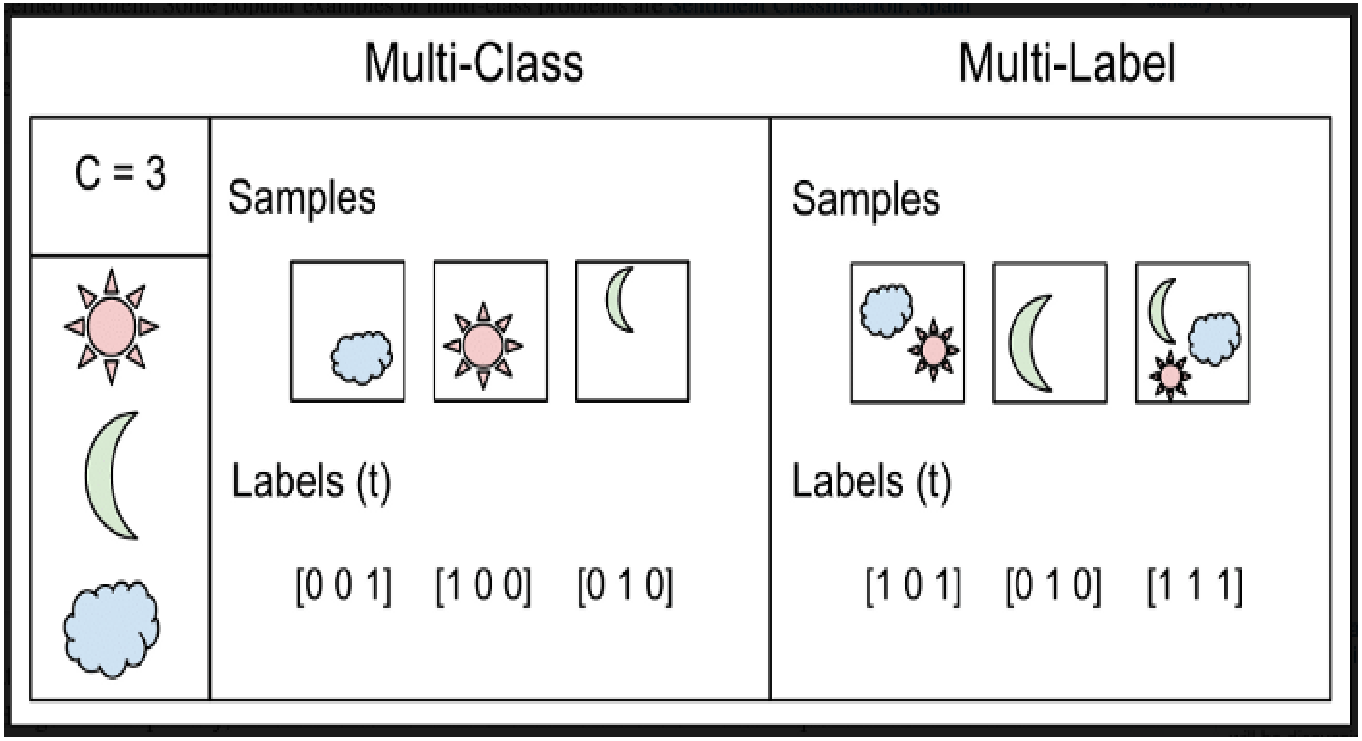
Multi-class Classification :-

Given a training set $\{ (X_i , Y_i) : X_i \in R^{n} , Y_i \in \{1,2,\ldots k\} , i = 1,2,.., l \}$, find a function f which can efficiently predict the label Y for an unseen $X \in R^{n}$.

Image Content Classification

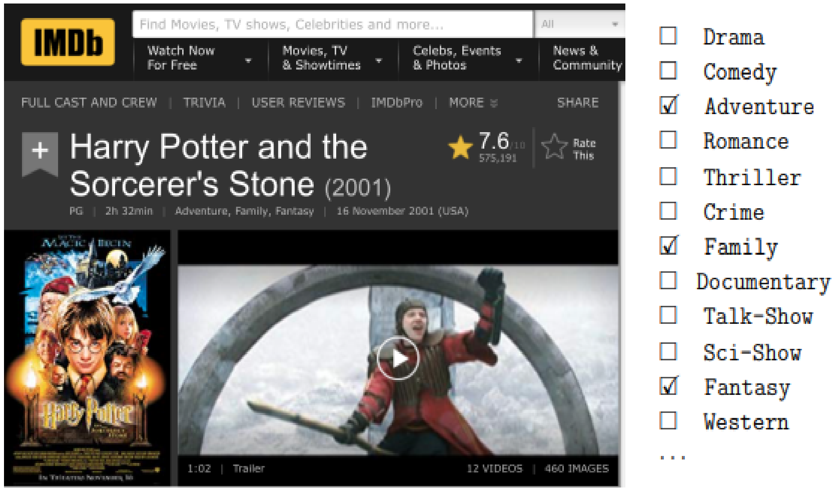


Multi- class and Multi label classification



Multi label classification

More Examples ??


IMDb
Find Movies, TV shows, Celebrities and more...
All
Watch Now For Free
Movies, TV & Showtimes
Celebs, Events & Photos
News & Community
FULL CAST AND CREW | TRIVIA | USER REVIEWS | IMDbPro | MORE
SHARE
Harry Potter and the Sorcerer's Stone (2001)
7.6/10 575,191
Rate This
PG | 2h 32min | Adventure, Family, Fantasy | 16 November 2001 (USA)
Harry Potter
1:02 | Trailer
12 VIDEOS | 460 IMAGES

- ☐ Drama
- ☐ Comedy
- ☑ Adventure
- ☐ Romance
- ☐ Thriller
- ☐ Crime
- ☑ Family
- ☐ Documentary
- ☐ Talk-Show
- ☐ Sci-Show
- ☑ Fantasy
- ☐ Western

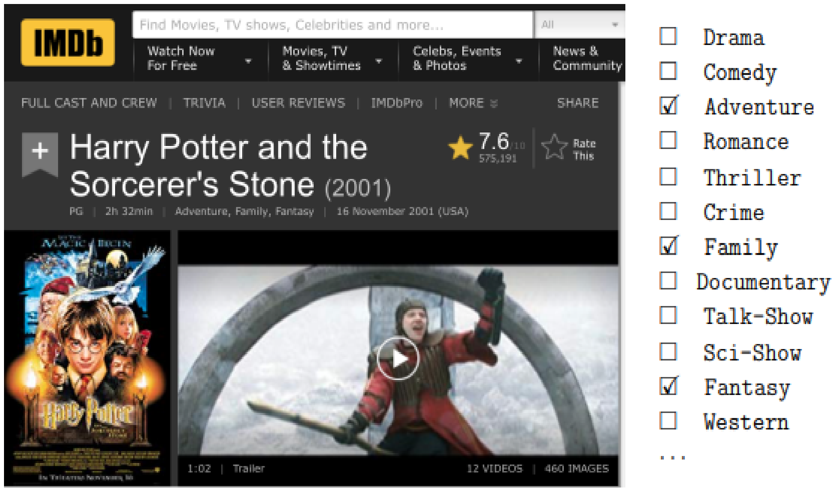
...

Figure 1.2: A screenshot of the IMDB webpage for the movie "Harry Potter and the Sorcerer's Stone". Genre tags associated with this movie = {Adventure, Family, Fantasy}. All genre tags available on IMDB = {Drama, Comedy, Romance, Thriller, Crime, Action, Horror, Adventure, Documentary, Mystery, Sci-Fi, Fantasy, Family, Biography, War, Animation, History, Music, Musical, Western, Short, Sport, Film-Noir, News, Adult, Talk-Show, Game-Show, Reality-TV}. From https://www.imdb.com/title/tt0241527/. Screenshot by author.

Multi label classification

Multi-label Classification :-

Given a training set $\{ (X_i , Y_i) : X_i \in R^n , Y_i \in \{0,1\}^L, i = 1,2,.., l \}$, where L is the number of labels, goal is to find a function f which can efficiently predict labels $Y \in \{0,1\}^L$ for an unseen $X \in R^n$.


IMDb
Find Movies, TV shows, Celebrities and more...
All
Watch Now For Free
Movies, TV & Showtimes
Celebs, Events & Photos
News & Community
FULL CAST AND CREW | TRIVIA | USER REVIEWS | IMDbPro | MORE
SHARE
Harry Potter and the Sorcerer's Stone (2001)
7.6/10
575,191
Rate This
PG | 2h 32min | Adventure, Family, Fantasy | 16 November 2001 (USA)
1:02 | Trailer
12 VIDEOS | 460 IMAGES

□ Drama
□ Comedy
☑ Adventure
□ Romance
□ Thriller
□ Crime
☑ Family
□ Documentary
□ Talk-Show
□ Sci-Show
☑ Fantasy
□ Western
...

Figure 1.2: A screenshot of the IMDB webpage for the movie "Harry Potter and the Sorcerer's Stone". Genre tags associated with this movie = {`Adventure`, `Family`, `Fantasy`}. All genre tags available on IMDB = {`Drama`, `Comedy`, `Romance`, `Thriller`, `Crime`, `Action`, `Horror`, `Adventure`, `Documentary`, `Mystery`, `Sci-Fi`, `Fantasy`, `Family`, `Biography`, `War`, `Animation`, `History`, `Music`, `Musical`, `Western`, `Short`, `Sport`, `Film-Noir`, `News`, `Adult`, `Talk-Show`, `Game-Show`, `Reality-TV`}. From `https://www.imdb.com/title/tt0241527/`. Screenshot by author.

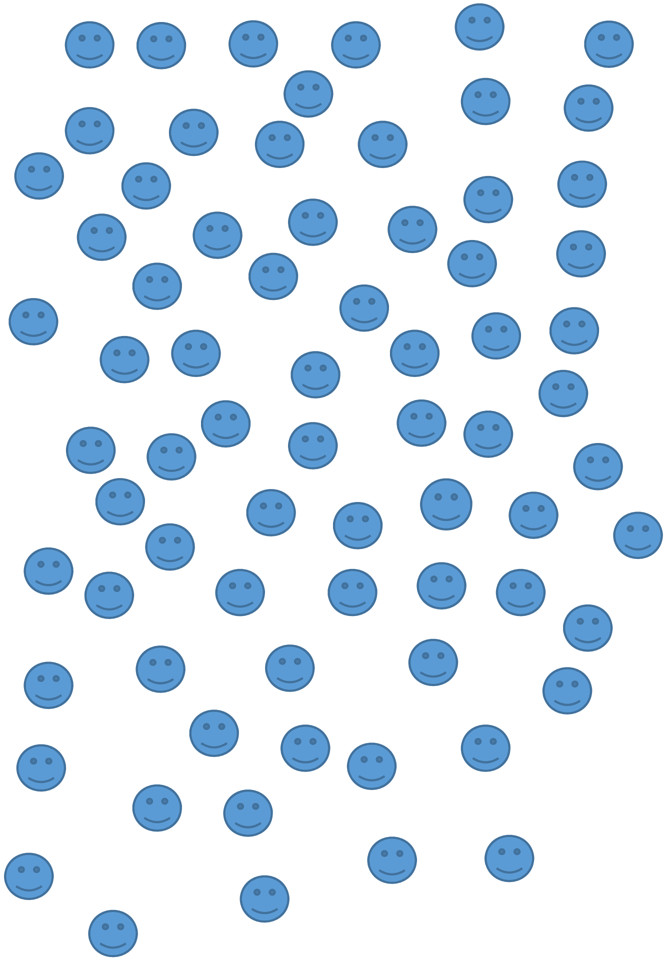
Regression Problem

Gender	Education	Seniority	Age	Work class	Income
0	3	3	34	1	78 K
1	2	2	65	0	89 K
0	1	4	25	0.	28 K
1	5	6	39	1	112 K
---	-----	---	------	------	--------
0	2	8	45	0	84 K
1	4	9	40	1	76 K
0	2	0	42	0	81 K

Given a training set $\{ (X_i , Y_i) : X_i \in R^n , Y_i \in R, i = 1,2,.., l \}$, find a function f which can efficiently approximate the relationship between independent variable X and Y for the prediction of response for an unseen test point $X_{test} \in R^{n}$.

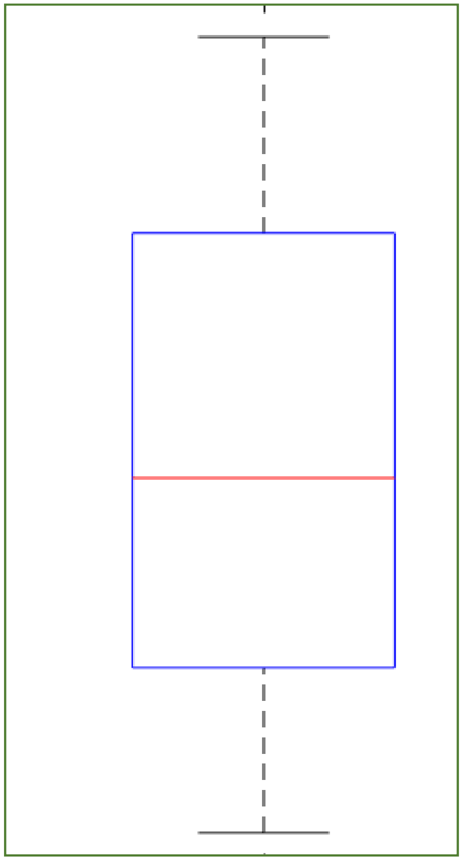
Credit Card Dataset

Income	Limit	Rating	Cards	Age	Balance
14.891	3606	283	2	34	333
106.025	6645	483	3	82	903
104.593	7075	514	4	71	580
148.924	9504	681	3	36	964
55.882	4897	357	2	68	331
80.18	8047	569	4	77	1151
20.996	3388	259	2	37	203
71.408	7114	512	2	87	872
15.125	3300	266	5	66	279
71.061	6819	491	3	41	1350
63.095	8117	589	4	30	1407



Weights in Kg.

55	84	62	88	90
73	105	57	62	54
53	102	83	81	54
101	86	68	78	71
50	60	53	57	60
72	48	104	107	77
77	67	104	69	52
96	70	60	71	51
53	107	108	82	48
60	68	99	46	47
56	87	93	87	74
81	64	94	57	90
57	69	86	96	50
106	96	77	73	74
65	78	78	98	97
87	69	98	79	94
74	50	60	105	55
99	80	110	50	74
52	108	45	96	98
102	50	71	62	97



Maximum = 110
Median = 74 Kg.
Minimum = 45.
25th Percentile :- 58.5
75th Percentile :- 94

55	84	62	88	90
73	105	57	62	54
53	102	83	81	54
101	86	68	78	71
50	60	53	57	60
72	48	104	107	77
77	67	104	69	52
96	70	60	71	51
53	107	108	82	48
60	68	99	46	47
56	87	93	87	74
81	64	94	57	90
57	69	86	96	50
106	96	77	73	74
65	78	78	98	97
87	69	98	79	94
74	50	60	105	55
99	80	110	50	74
52	108	45	96	98
102	50	71	62	97

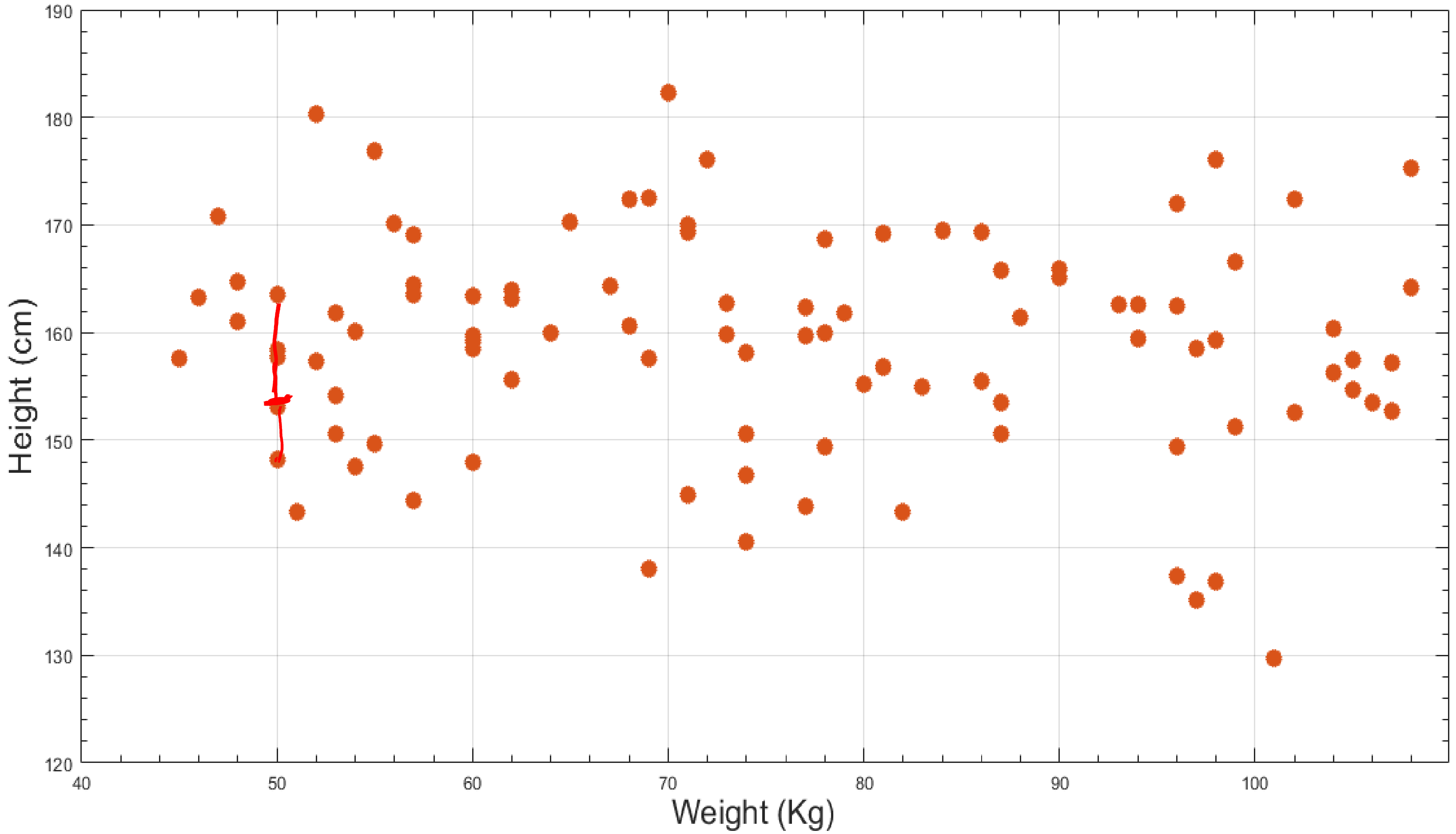
Regression Problem

Heights in Cm. (Y)

161.8	157.5	149.3	170.0	169.1
149.7	149.4	169.3	143.4	165.9
169.5	176.0	163.5	154.1	163.5
163.1	172.3	159.7	157.2	172.5
161.4	157.7	161.8	164.2	169.3
165.2	144.9	144.3	143.3	162.4
162.6	155.6	159.2	164.7	153.1
150.6	158.4	176.0	147.9	153.5
158.4	162.8	161.0	160.7	171.9
158.5	157.4	160.4	166.5	143.9
154.7	164.4	152.7	163.3	159.8
176.8	163.9	159.7	170.8	140.5
151.2	147.5	162.3	170.1	170.2
155.2	150.5	164.3	153.5	168.6
152.9	152.6	156.3	162.6	160.0
148.3	154.9	157.6	150.6	159.3
158.1	156.8	180.2	146.8	135.1
157.3	160.1	137.4	169.2	165.8
175.3	129.7	182.3	160.0	138.1
155.4	172.4	163.4	159.5	136.8

Weights in Kg. (X)

55	84	62	88	90
73	105	57	62	54
53	102	83	81	54
101	86	68	78	71
50	60	53	57	60
72	48	104	107	77
77	67	104	69	52
96	70	60	71	51
53	107	108	82	48
60	68	99	46	47
56	87	93	87	74
81	64	94	57	90
57	69	86	96	50
106	96	77	73	74
65	78	78	98	97
87	69	98	79	94
74	50	60	105	55
99	80	110	50	74
52	108	45	96	98
102	50	71	62	97


190
180
170
160
150
140
130
120
40
50
60
70
80
90
100
Weight (Kg)
Height (cm)

Regression Problem

Identically and Independently distributed

Regression Assumptions:-

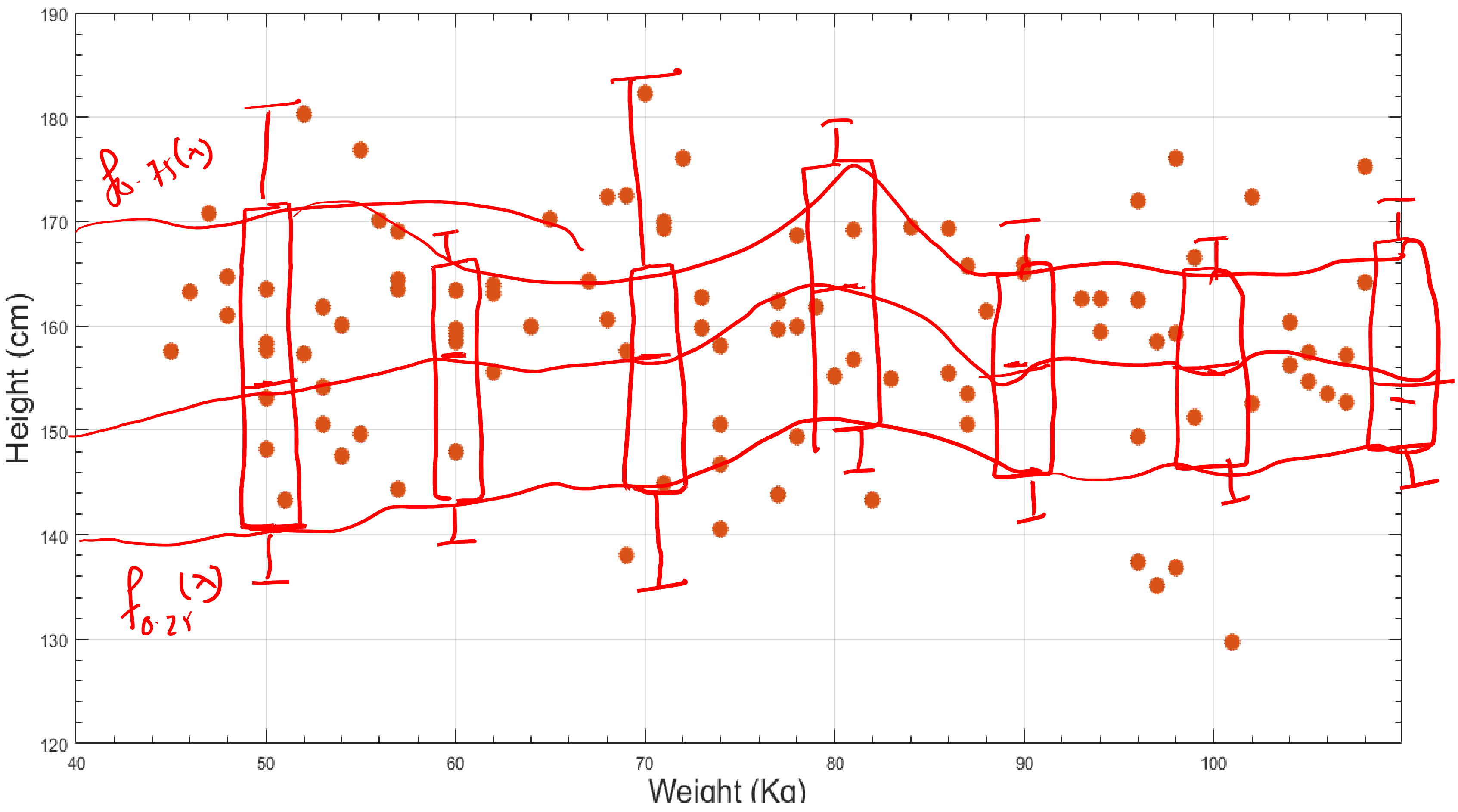
(i) x_i, y_i are iid random variables.

(ii) $$y_i = \underbrace{f_0(x_i)}_{\uparrow \atop E(Y/x_i)} + \epsilon_i$$

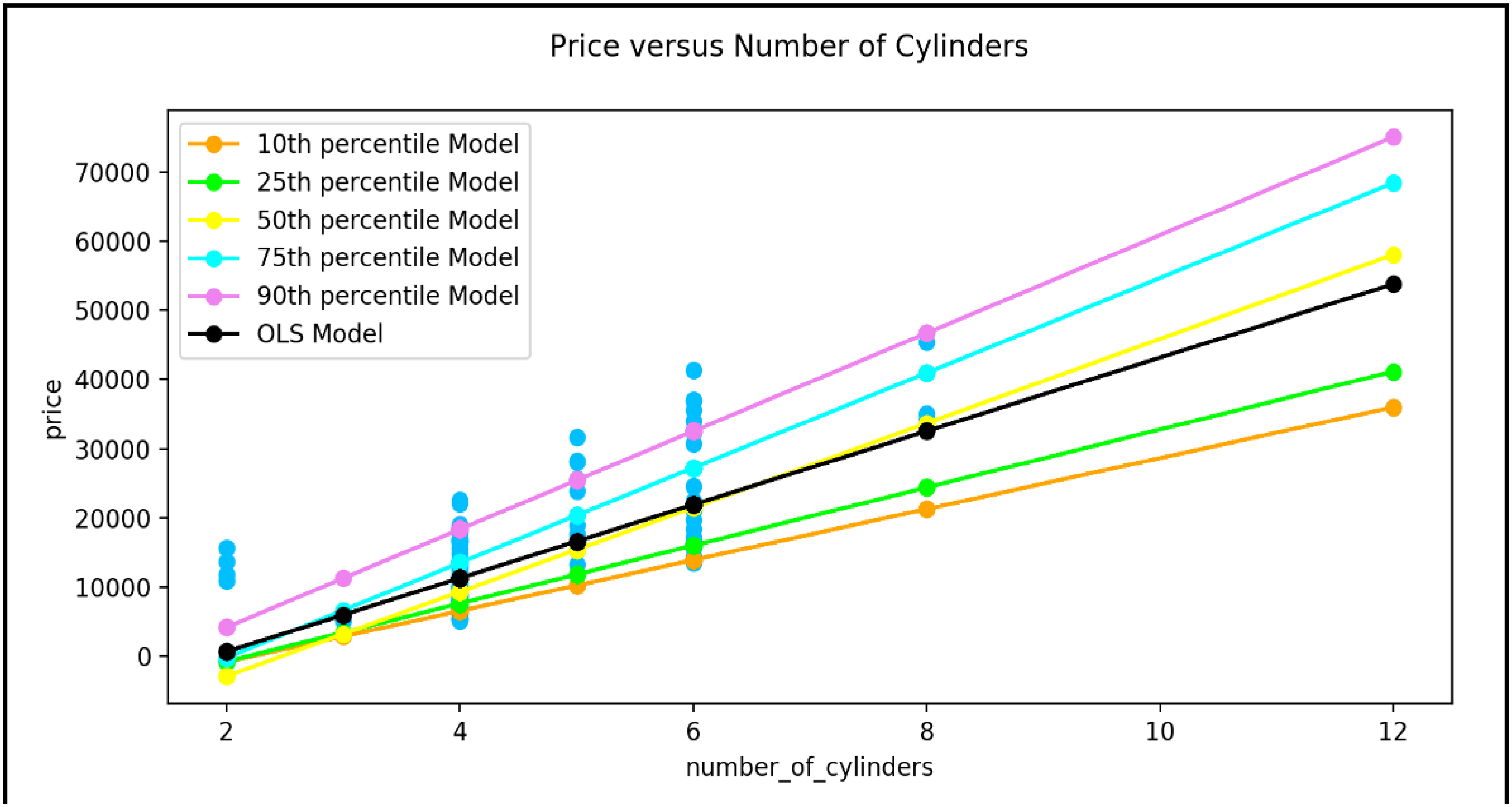
(iii) $$E(\epsilon_i) = \underline{0}$$

Given a training set $\{ (X_i, Y_i) : X_i \in R^n, Y_i \in R, i = 1,2,.., l \}$, find a function f which can efficiently approximate the relationship between independent variable X and Y for the prediction of response for an unseen test point $X_{test} \in R^n$.

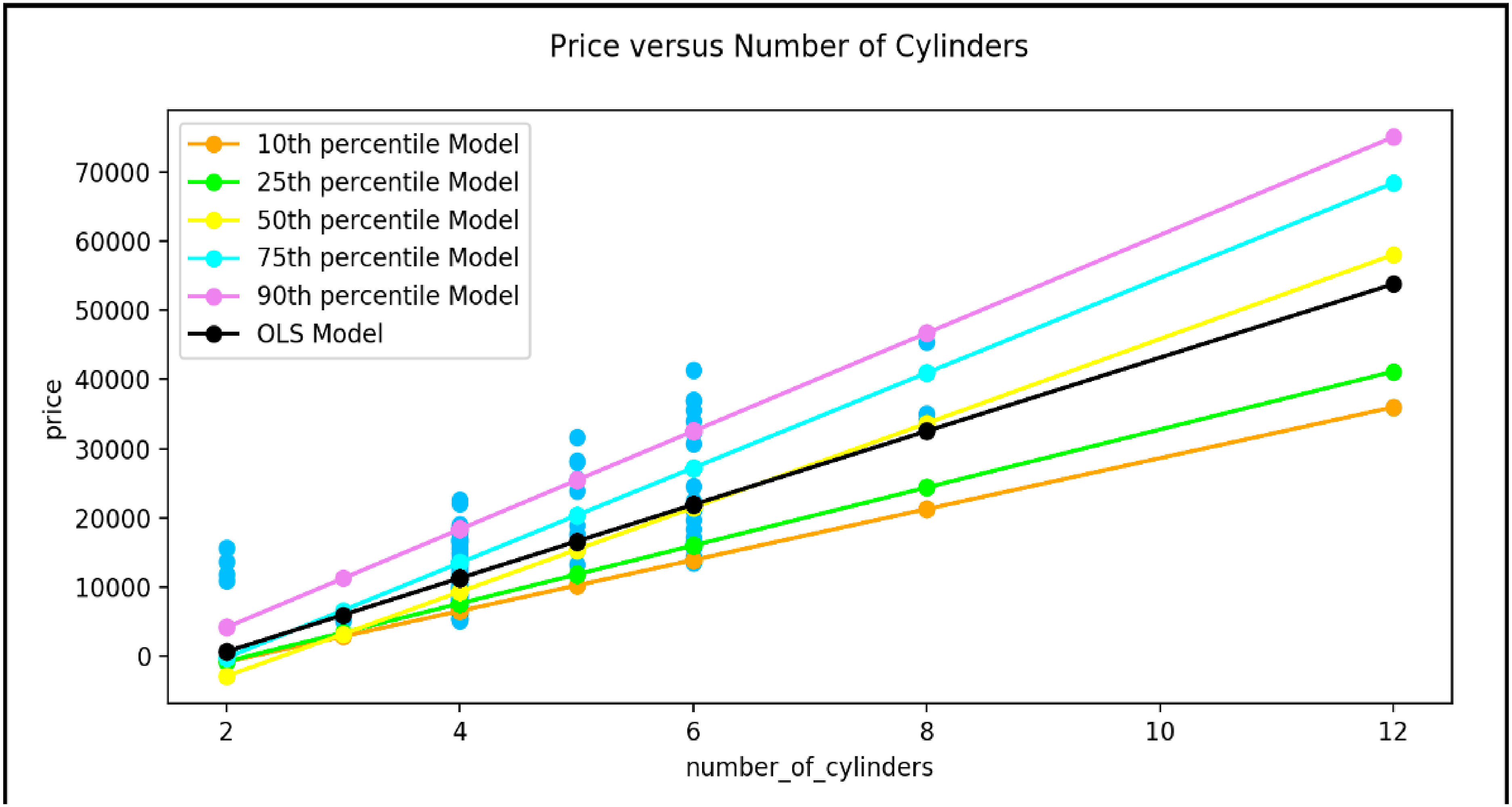
Beyond Mean Regression Problem



Quantile Regression Problem



More Examples



Quantile Regression Problem

For given traing set $T = \{(x_i, y_i) : x_i \in R^n, \; y_i \in R, \; i = 1, 2, \dots l\}$, and quantile $\tau \in (0,1)$, the QR model

estimate a function $u(x)$ such that it is infimum of

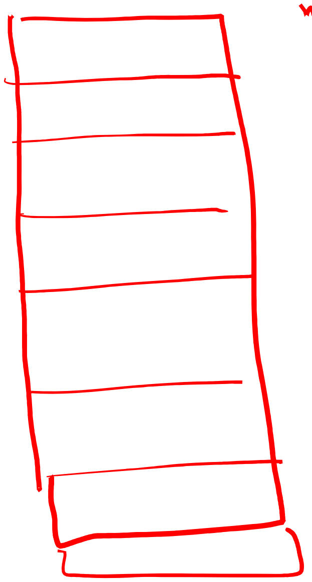
the set of function satisfying $\quad P(y \leq u_i(x) \mid x) = \tau.$

Multi-output Regression Problem

Y_1

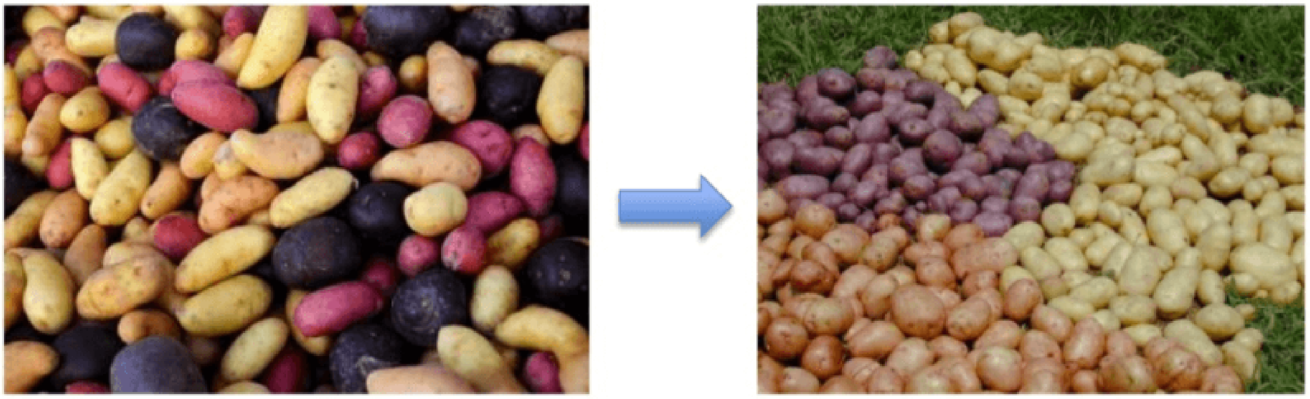
Income	Limit	Rating	Cards	Age	Balance
14.891	3606	283	2	34	333
106.025	6645	483	3	82	903
104.593	7075	514	4	71	580
148.924	9504	681	3	36	964
55.882	4897	357	2	68	331
80.18	8047	569	4	77	1151
20.996	3388	259	2	37	203
71.408	7114	512	2	87	872
15.125	3300	266	5	66	279
71.061	6819	491	3	41	1350
63.095	8117	589	4	30	1407

Y_2 Investenet in mutual funds

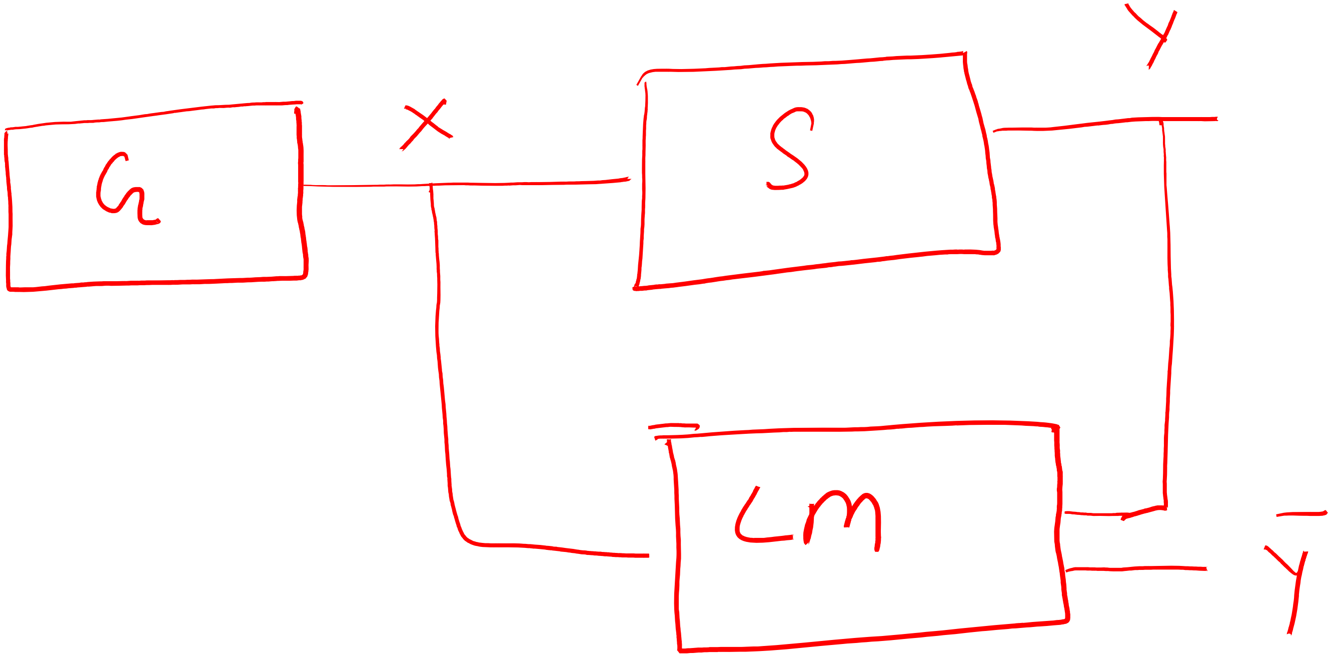


Unsupervised Learning

Clustering

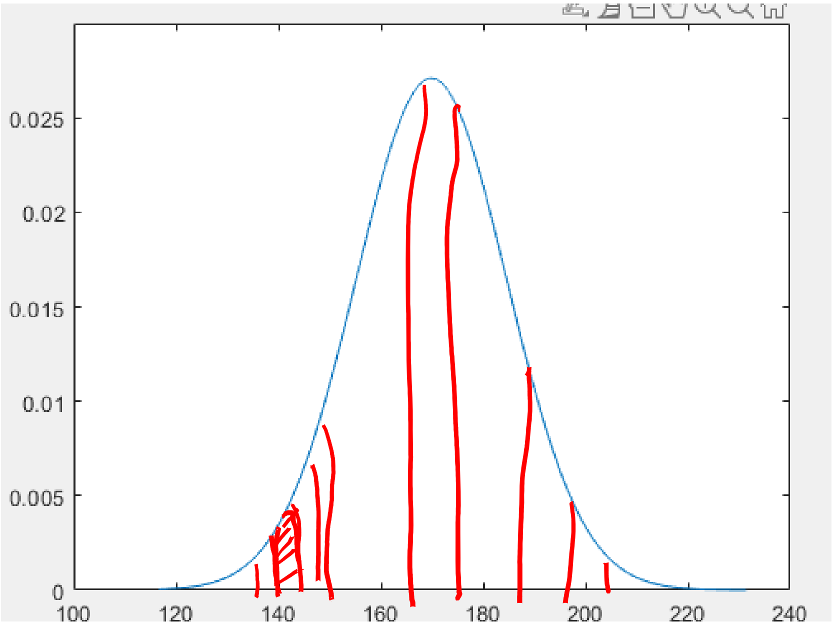


Unsupervised Learning



$$\int_{140}^{142} f(x)\,dx$$

$$\sum_{x} f(x)\,\delta x$$


0.025
0.02
0.015
0.01
0.005
0
100
120
140
160
180
200
220
240