1) Roll no. 18ME 10060

(i) Random seed : 10060

Using this we get

$X = [10, -50, 18, -26, -46, 19, 11, 14, 13, -1]$

Followed by this we generate

Y as given by formula

$Y = x^2 + 7x + 4.$

X	Y	X	Y
10	174	19	498
-50	2154	11	~~298~~ 202
18	454	14	298
-26	498	13	264
-46	1798	-1	-2

So 1 to 8 points are training set.

(ii) mean of Y in training set

$$y_0 = \frac{6076}{8} = 759.5$$

$\boxed{R_i = Y_i - y_0}$

Sr.	~~Y~~	X	R
1	174	10	-585.5
2	2154	-50	1394
3	454	18	-305.5
4	498	-26	-261.5
5	1798	-46	1038.5
6	498	19	-261.5
7	202	11	-557.5
8	298	14	-461.5

(iii) Linear regression X_i vs R_i. | convert X to (8, 2)

We will use ordinary least square method for calculating w_1 & b.

$b = w_0$

$$\bar{w} = (X^T X)^{-1} (X^T y)$$

$\boxed{\text{here } y \text{ is } R_i.}$

in python this'll be

```
a = np.linalg.inv(np.dot(X^T, X))
b = np.dot(X.T, y)
w = np.dot(a, b)
print(w)
```

$\overline{w} = [-143.48, -22.957] \equiv [w_0, w_1]$

iv) Now the next step is

$$y_{1i} = y_0 + \alpha 1 * (w1 * x_i + b1)$$

we need $\alpha 1$ s.t.:
such that $\sum_i (y_{1i} - y_i)^2$ is least

So, this could be solve using a simple quadratic equation

$$\frac{\partial L}{\partial \alpha 1} = 0 \qquad L = \sum_i (y_{1i} - y_i)^2$$

$$\frac{\partial L}{\partial \alpha 1} = \frac{\partial (3205095.7\alpha 1^2 - 6410191.48\alpha 1 + 4119806)}{\partial \alpha 1}$$

$$\boxed{\alpha 1 = +0.999}$$

The value of L at $\alpha 1 = +0.999$ is 914711.

Hence we calculate new residuals

$$\boxed{R_i = y_i - y_{1i}}$$

Sr. No.	y_i	R_i'
1	174	-212.82
2	2154	391.14
3	454	250.65
4	498	-714.46
5	1798	126.87
6	498	317.58
7	202	-161.89
8	298	2.91

(V) Now we again fit a linear model for x_i Vs R_i'

Iter 1:

This time

$$\overline{w_2} = [-0.143, -0.023] \equiv [w_0, w_1]$$

Next we define

$$y_{2i} = y_0 + \alpha_1^* (w_1^* x_i + b_1) + \alpha_2^* (w_2^* x_i + b_2)$$

again α_2 s.t. $\sum_i (y_{2i} - y_i)^2$ is min.

$$L = 3.2\alpha_2^2 - 6.410\alpha_2 + 914713$$

$$\frac{dL}{d\alpha} = 2\alpha_2 (3.2051) - 6.41019 = 0$$

$$\boxed{\alpha_2 = 0.999}$$

JUNE 2020

Su	Mo	Tu	We	Th	Fr	Sa
	1	2	3	4	5	6
	8	9	10	11	12	13
	15	16	17	18	19	20
	22	23	24	25	26	27
	29	30				

Hence we calculate new Residuals R_i''

Sr. No.	y_i	R_i''
1	174	-212.45
2	2154	390.13
3	454	251.21
4	498	-714.9
5	1798	125.96
6	498	318.16
7	202	-161.49
8	298	3.38

Iter 2:

Linear model for X_i vs R_i''

$$\overline{W3} = [-0.000143, -2.2957 \times 10^{-5}]$$

We now define

$$Y3i = Y0 + \alpha 1^* (W1^* X_i + b1) + \alpha 2 (\overline{W2}^* X_i) + \alpha 3 (\overline{W3}^* X_i)$$

Again we solve for $\alpha 3$ s.t. Loss is minim

$$\boxed{\alpha 3 = 0.999}$$

$$3.2051 \times 10^{-6} \alpha 3^2 - 6.410 \alpha 3 + 914710.25 = 0$$

(vi) Hence finally our set of α is

$(\alpha_1, \alpha_2, \alpha_3) \equiv (0.999, 0.999, 0.999)$

$\overline{w} \equiv [-143.48, -22.957]$

$\overline{w_2} \equiv [-0.143, -0.023]$

$\overline{w_3} \equiv [-0.000143, -2.2957 \times 10^{-5}]$

Now my test set $X_t = [13, -1]$
$y_t = [264, -2]$

$$y_{pred} = y_0 + 0.999 (w_1^T x_t + b_1 + w_2^T x_t + b_2 + w_3^T x_t + b_3)$$

$$= 759.5 + 0.999 \left([143.48, -22.957] \begin{bmatrix} 1 & 1 \\ 13 & -1 \end{bmatrix} + [-0.143 \; -0.023] \begin{bmatrix} 1 & 1 \\ 13 & -1 \end{bmatrix} + [-0.000143 \; -2.2957\times10^{-5}] \begin{bmatrix} 1 & 1 \\ 13 & -1 \end{bmatrix} \right)$$

$$= [759 \quad 759] + [y_{p1} \quad y_{p2}]$$

$$= [317 \quad 638]$$

16

Tuesday

Hence the predicted Values of y for test set is [317, 638]

2] · K-means with $K = 2$, $N = 11$

Lets consider the dataset as seen on the plot before. (Fig 1)

$$\begin{cases} x_1 : 1, 2, 3, 1, 3, 5, 6, 7, 5, 7, 5 \\ x_2 : 1, 2, 3, 3, 1, 5, 6, 7, 7, 5, 4 \end{cases}$$

This is our unlabeled dataset, we wanna try to make two clusters from it.

Let us first randomly initilize with centres $\bar{X}_{C_1} := (1, 1)$ and $\bar{X}_{C_2} := (1, 2)$

Iter 1 :

x_1	x_2	category ~~Label~~
1	1	C_1
2	2	C_1
3	3	C_1
1	3	C_1
3	1	C_1
5	5	C_2
6	6	C_2
7	7	C_2
5	7	C_2.
7	5	C_2
5	4	C_2

$\bar{X}_{C_1} = (2, 1)$

$\bar{X}_{C_2} = (4.55, 4.66)$

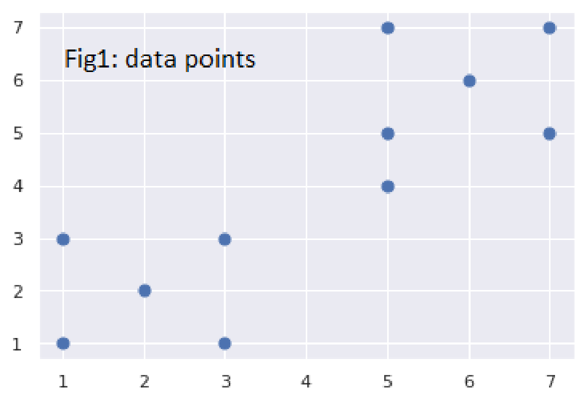
Fig1: data points
7
6
5
4
3
2
1
1
2
3
4
5
6
7

After one iteratⁿ, cluster centers are updated as follows

$\bar{X}_{ci}$:= mean of all points under the cluster i.

Iter 2 : new centers

x_1	x_2	category label
1	1	C_1
2	2	C_1
3	3	C_1
1	3	C_1
3	1	C_1
6	6	C_2
7	6	C_2
5	5	C_2
7	5	C_2
5	7	C_2
5	4	C_2

$\bar{X}_{C_1} = (2, 2)$

$\bar{X}_{C_2} = (5.83, 5.66)$

Iter 3 : new centers

x_1	x_2	category label
1	1	C_1
2	2	C_1
3	3	C_1
1	3	C_1
3	1	C_1
5	5	C_2
6	6	C_2
7	7	C_2
5	7	C_2
7	5	C_2
5	4	C_2

$\bar{X}_{C_1} = (2, 2)$

$\bar{X}_{C_2} = (5.83, 5.66)$

27 Sunday

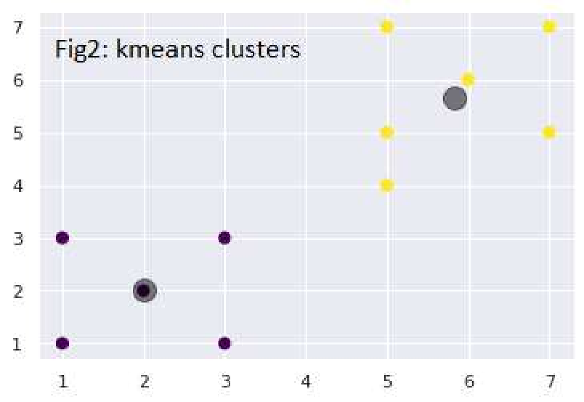
Fig2: kmeans clusters
7
6
5
4
3
2
1
1
2
3
4
5
6
7

Hence, the process terminates as the cluster centre don't vary with increased iteration. Also we get the desired clusters based on geometric interpretation. (

Now lets try again with
$\overline{X}_{C_1} := (6, 4)$ and $\overline{X}_{C_2} := (2, 6)$

X_1	X_2	category ~~label~~
1	1	C_1
2	2	C_1
3	3	C_1
1	3	C_2
3	1	C_1
5	5	C_1
6	6	C_1
7	7	C_1
5	7	C_2
7	5	C_1
5	4	C_1

$\overline{X}_{C_1} := (4.33, 3.78)$

$\overline{X}_{C_2} := (3, 5)$

Iter 2 :

x_1	x_2	category ~~label~~
1	1	C_1
2	2	C_1
3	3	C_1
3	1	C_1
1	3	C_2
5	5	C_1
6	6	C_1
7	7	C_1
5	7	C_2
7	5	C_1
5	4	C_1

$\bar{x}_{c_1} := (4.33, 3.78)$

$\bar{x}_{c_2} := (3, 5)$

As the cluster centres don't deviate with the next iteration we stop here.
Refer to fig 3 for representation
Clearly this shows us how K-means fails miserably because of its limitation to initialization.

This can be solved using Kmeans++.

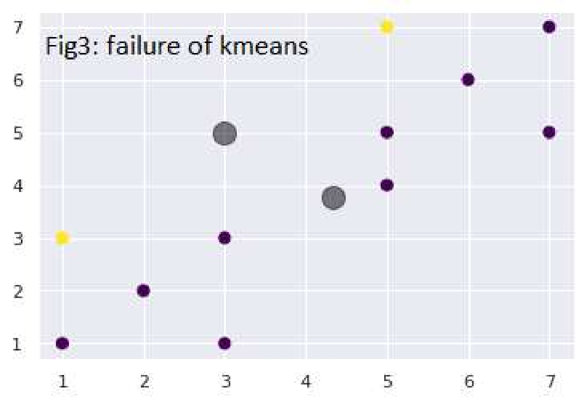
Fig3: failure of kmeans
7
6
5
4
3
2
1
1
2
3
4
5
6
7

In Kmeans++, we compute the distance of all the point from the nearest centroid (t-1). The one which has largest distance has highest probability for being new center.

So lets begin with $\bar{X}_{C_1}$ as (6, 1).
Let $\bar{a}_i$ present (x_i, y_i)

X_1	X_2	$\|C_1 - a_i\|_2$	Prob	$p \cdot X_1$	$p \cdot X_2$
1	1	5	0.100	0.1	0.1
2	2	4.123	0.083	0.166	0.166
3	3	3.605	0.072	0.216	0.216
3	1	3	0.060	0.180	0.060
1	3	5.38	0.108	0.108	0.324
5	5	4.123	0.083	0.415	0.415
6	6	5	0.100	0.6	0.6
7	7	6.08	0.1224	0.857	0.857
5	7	6.08	0.1224	0.612	0.857
7	5	4.123	0.083	0.581	0.415
5	4	3.16	0.064	0.32	0.256
			$\Sigma = 1$	4.155	4.27

Hence the other cluster center
$\bar{X}_{C_2} := (4.155, 4.27)$

Iter 1

x_1	x_2 ~~x_2~~	category
1	1	C_1
2	2	C_1
3	3	C_2
1	3	C_1
3	1	C_1
5	5	C_2
6	6	C_2
7	7	C_2
5	7	C_2
7	5	C_2
5	4	C_2

$\overline{x}_{C_1} := (3, 1)$

$\overline{x}_{C_2} := (4.2, 4.3)$

Iter 2

x_1	x_2	category
1	1	C_1
2	2	C_1
3	3	C_1
1	3	C_1
3	1	C_1
5	5	C_2
6	6	C_2
7	7	C_2
5	7	C_2
7	5	C_2
5	4	C_2

$\overline{x}_{C_1} := (1.75, 1.75)$

$\overline{x}_{C_2} := (5.43, 5.28)$

Monday

Iter 3

x_1	x_2	Category
1	1	C_1
2	2	C_1
3	3	C_1
1	3	C_1
3	1	C_1
5	5	C_2
6	6	C_2
7	7	C_2
5	7	C_2
7	5	C_2
5	4	C_2

$\bar{X}C_1 := (2,2)$

$\bar{X}C_2 := (5.83, 5.6$

Iter 4

x_1	x_2	Category
1	1	C_1
2	2	C_1
3	3	C_1
1	3	C_1
3	1	C_1
5	5	C_2
6	6	C_2
7	7	C_2
5	7	C_2
7	5	C_2
5	4	C_2

$\bar{X}C_1 := (2,2)$

$\bar{X}C_2 := (5.83, 5.67)$

Hence, we stop as the center doesn't dievate

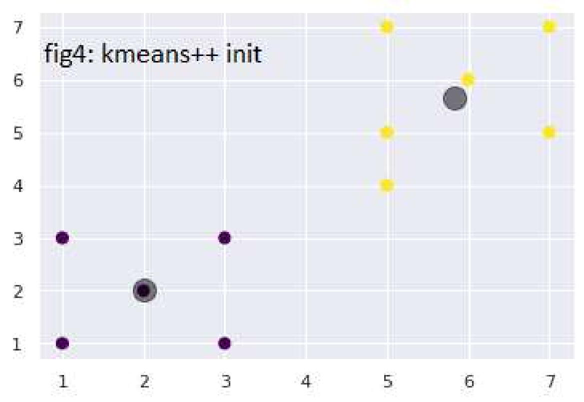
fig4: kmeans++ init
1
2
3
4
5
6
7
1
2
3
4
5
6
7

③ Ans Maximum - likelihood estimation:

(i) Geometric distribution:

Let $X_1, X_2, X_3 \ldots X_n$ be a random sample from the geometric distribution with p.d.f.

$$f(x, p) = (1-p)^{x-1} p, \quad x = 1, 2, 3, \ldots$$

The likelihood function is given by

$$L(p) = (1-p)^{x_1-1} p \, (1-p)^{x_2-1} p \ldots = p^n (1-p)^{\left(\sum_{i=1}^{n} x_i - n\right)}$$

Take log

$$\ln[L(p)] = n \ln p + \left(\sum_{i=1}^{n} x_i - n\right) \ln(1-p)$$

$$\frac{d}{dp} \ln[L(p)] = \frac{n}{p} - \frac{\left(\sum_{i=1}^{n} x_i - n\right)}{(1-p)} = 0$$

$$p = \frac{n}{\sum_{i=1}^{n} x_i}$$

$\therefore$ The maximum liklihood estimator of P is

$$\boxed{p = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}}$$

9 (ii) Binomial Distribution :

10

pmf is K is no. of successes in n trails (1)

11

$$f(x) = {}^nC_x \, p^x \cdot (1-p)^{n-x}$$

12

Liklihood function

1

$$L(p) = \prod_{i=1}^{n_1} \left({}^nC_{x_i} \, p^{x_i} \, (1-p)^{n-x_i} \right)$$

2

3 $$L(p) = \left[\prod_{i=1}^{n_1} \left(\frac{n!}{(x_i)!\,(n-x_i)!} \right) \right] p^{\sum_i^{n_1} x_i} \, (1-p)^{n n_1 - \sum_{i=1}^{n} x_i}$$

4 Z

5 $$\ln(L(p)) = \sum_{i=1}^{n_1} x_i \ln(p) + \left(n n_1 - \sum_{i=1}^{n_1} x_i\right) \ln(1-p) + \ln(Z)$$

6

Taking derivative and equating to zero.

7

$$\frac{d \ln(L(p))}{dp} = \frac{1}{p} \sum x_i - \frac{1}{(1-p)} \left(n_1 n - \sum_{i=1}^{n_1} x_i\right) = 0$$

$$p = \frac{\sum_{i=1}^{n_1} x_i}{n_1 n}$$

$\cancel{p / x}$ ~~Refer K from (1)~~

$$\boxed{\hat{p} = \frac{\bar{x}}{n}}$$

⇒ This is the MLE for binomial distribution

③ Ans Now μ is known but σ^2 unknown

Hence we'll try to work out via precision $\lambda = \frac{1}{\sigma^2}$, cause it is better to work with parameter in numerator

$$\prod_{i=1}^{N} \text{prob}(x_i / \mu, \lambda) = \frac{\lambda^{N/2}}{(2\pi)^{N/2}} \exp\left\{-\frac{\lambda}{2}\sum_{i}^{N}(x_n - \mu)^2\right\}$$

where $x_1, x_2 \ldots x_N$: Data and λ is our parameter

Also we know that

$$\text{post}(\lambda/x) \propto \prod \text{prob} \cdot \text{prior}(\lambda)$$

It is also mentioned that the prior has inverse-gamma function (but this'll change)

$$F(\lambda; \mu_0, \sigma_0) = \frac{\sigma^{-(\mu_0+1)} e^{-\frac{1}{\sigma\sigma_0}}}{\Gamma(\mu_0)(\sigma_0)^{\mu_0}}$$

Now $\Gamma(\mu_0)$ is a constant so it's not a problem

$post(\lambda/D) \propto prob(D/\lambda, u) \cdot prior(\lambda)$

Also we know that if $\boxed{(Cond - 1)}$
$X \sim Gamma(\alpha, B)$ then $\frac{1}{X} \sim Inv\text{-}gamma(\alpha, \beta)$
↳ ①

So technically we had parameter variance is inverse gamma so by ① we deduce parameter precision has gamma prior distributⁿ.

Hence the prior is now gamma(a_0, b_0)

$$\Rightarrow post(\lambda/D) \propto \lambda^{N/2} \exp\left(-\lambda \sum_i^N (x_i - u)^2\right)\left(\lambda^{a_0-1} \exp(-b_0\lambda)\right)$$

~~Note the~~ term $b_0^{a_0}$ & $\Gamma(a_0)$ are just normalizing constants.

$$\Rightarrow post(\lambda/D) = \lambda^{N/2 + a_0 - 1} \exp\left\{-\lambda\left(b_0 + \sum (x_n - u)^2\right)\right\}$$

Now we have to make manipulations for a_0' and b_0' s.t. post(λ/D) also has same distribution

From MLE

$$a_0' = \frac{N}{2} + a_0 \qquad b_0' = b_0 + \frac{N}{2}\sigma^2_{ML}$$

Hence

$$\text{post}(\lambda / D) = \text{gamma}\left(\lambda ; \frac{N}{2} + a_0 , b_0 + \frac{N}{2}\sigma^2_{ML}\right)$$

Now we also now $\lambda = \frac{1}{\sigma^2} \Rightarrow \sigma^2 = \frac{1}{\lambda}$

Using condition $\boxed{\text{cond} - 1}$ we can state

$$\text{post}(\sigma^2 / D) = \text{inv-gamma}\left(\sigma^2 ; \frac{N}{2} + a_0 , \left(b + \frac{N}{2}\sigma^2_{ML}\right)\right)$$

Hence the posterior distribution of the variance parameter for 1-d (prior ≡ inverse-gamma) for Gaussian is known.

$$\sigma^2_{ML} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \quad \text{as } N \to 0$$

the value as a_0' & b_0' will become a_0, a_0 i.e. only dependent on prior.

$\pi(x) = $

OCTOBER 2020

26 Monday

300-066 | Week 44

27 Saturday

179-187 | Week 26

③ Ans → Another approach : directly through inverse gamma

Data : $\{x_1, \ldots x_N\}$

Model : $X_i \sim N(\mu, \sigma)$

$\sigma \in R$, so prior$(\sigma) =$ ~~gatam~~$(\alpha_0, \beta_0) = F(\sigma; \alpha_0, \beta_0)$

So clearly the assumption for this Bayesian estimate is that μ is known.

$\therefore$

$$\text{posterior}(\sigma/X) = \prod_i^N \text{prob}\left(X_i = x_i/\sigma\right) \cdot \text{prior}(\sigma)$$

$$= \frac{1}{\sigma^N} \exp\left(-\frac{\sum_i^N (x_i - \mu)^2}{2\sigma^2}\right) \cdot \text{prior}(\sigma)$$

$$= \cancel{\frac{1}{\sigma^N} \exp\left(-\frac{\sum_i^N (x_i - \mu)^2}{2\sigma^2}\right)} \cdot \underbrace{\cancel{\frac{(\sigma)^{\alpha_0 - 1} (1+\sigma)^{-\alpha_0 - \beta_0}}{B(\alpha_0, \beta_0)}}}_{1}$$

28 Sunday

~~Now lets keep our focus of simplifying the term 1.~~

$$\Rightarrow \cancel{\frac{(\sigma)^{u_0-1}(1+\sigma)^{-u_0-\sigma_0}}{\Gamma(u_0)\,\Gamma(\sigma_0)/\Gamma(u_0+\sigma_0)}}$$

$$= \frac{1}{\sigma^N}\exp\left(-\frac{\sum_i^N (x_i-u)^2}{2\sigma^2}\right)\cdot\frac{\sigma^{-(u_0+1)}\,e^{-\frac{1}{\sigma_0\sigma}}}{\Gamma(u_0)\,(\sigma_0)^{u_0}}$$

$$F(\sigma;u_0,\sigma_0) = \frac{\sigma^{-(u_0+1)}\,e^{-\frac{1}{\sigma\sigma_0}}}{\Gamma(u_0)\,(\sigma_0)^{u_0}}$$

$\Gamma(u_0)$ is gamma function of u_0 which'll be a constant.

$$= \frac{1}{\sigma^{(N+u_0+1)}}\,\frac{1}{\Gamma(u_0)(\sigma_0)^{u_0}}\exp\left(-\frac{\sum_{i=1}^{N}(x_i-u)^2}{2\sigma^2}-\frac{1}{\sigma_0\sigma}\right)$$

So the prior and posterior have same distribution but with different parameter so we try to convert the above expression into inverse-gamma.

$$F(x;\alpha,B) = \frac{(x)^{-(\alpha+1)}\,e^{-\frac{1}{Bx}}}{\Gamma(\alpha)\,B^{\alpha}}$$

$$\Rightarrow \frac{K \quad \sigma_0^{N}}{\sigma^{(N+u_0+1)} \Gamma(N+u_0) \, (\sigma_0)^{(N+u_0)}} \exp\left(\frac{-N \sigma_{ML}^{*}}{2\sigma^2} - \right) \sigma_0$$

where

K is $(N+u_0-1) \times (N+u_0-2) * \cdots * (u_0)$

$$\Rightarrow \frac{K \quad \sigma_0^{N}}{\sigma^{(N+u_0+1)} \Gamma(N+u_0) (\sigma_0)^{(N+u_0)}} \exp\left(-\frac{1}{\sigma}\left(\frac{N}{2} + \frac{1}{\sigma_0}\right)\right)$$

$$\Rightarrow \frac{K \sigma_0^{N}}{\sigma^{(N+u_0+1)} \Gamma(N+u_0) \, \sigma_0^{(N+u_0)}} \exp\left(-\frac{1}{\sigma \left(\frac{2\sigma_0}{(N\sigma_0+2)}\right)}\right)$$

$$\text{post}(\sigma^2, a_0', b_0') \equiv \text{inv-gamm}\left(\sigma, \frac{N}{2} + u_0 + 1, \frac{2\sigma_0}{N\sigma_0+2}\right)$$