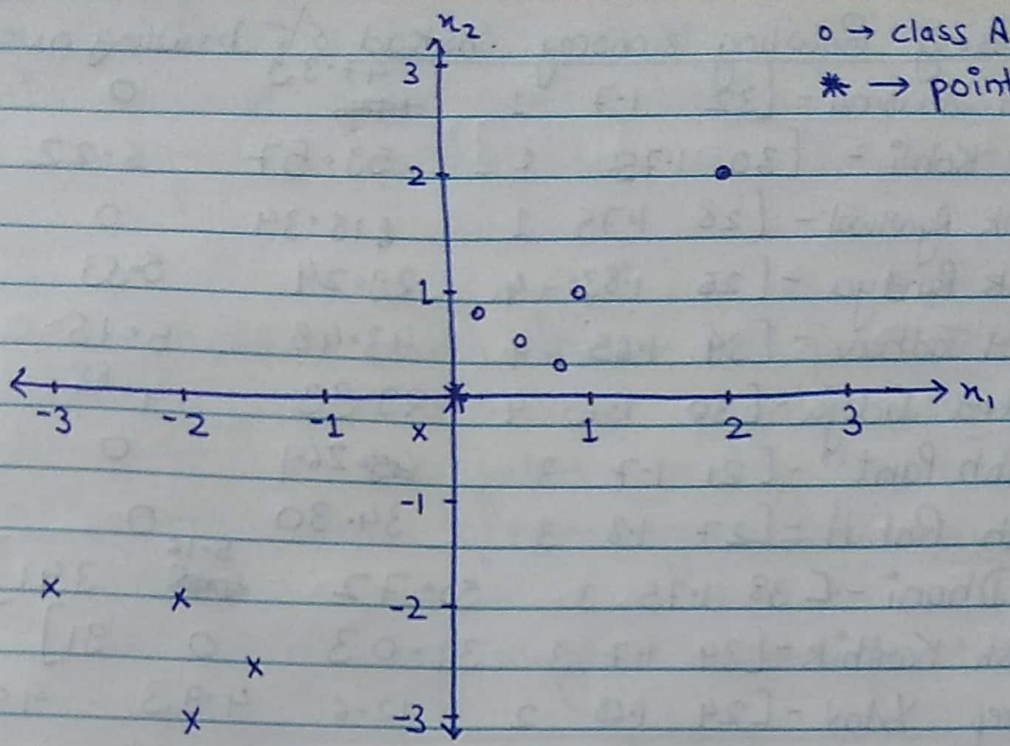
Q1



$o \rightarrow$ class A ; $\times \rightarrow$ class B

$* \rightarrow$ point q

→ K=1

Now when k=1, the nearest neighbour is point $[-0.25 \ -0.25]^T$ as it can be seen from the graph. Since $[-0.25 \ -0.25]^T$ Belongs to class B, thus q also belongs to class B

$\therefore q \in B$

→ K=3

Now when K=3, the nearest 3 neighbours are $[-0.25 \ -0.25]^T$, $[0.5 \ 0.5]^T$ and $[0.75 \ 0.25]^T$ $[0.25 \ 0.75]^T$ (both are equidistant) Since out of these 4 points, 3 of them belong to class A, so the point q belongs to class A

$\therefore q \in A$

Q2. * Using Bowling economy instead of bowling average

Rohit Sharma - [32 1.7 1 47.39 0 206]
Virat Kohli - [30 1.75 1 59.57 6.22 227]
Mayank Agarwal - [28 1.75 1 18.34 0 77]
Hardik Pandya - [25 1.83 4 29.24 5.53 45]
Kedar Jadhav - [34 1.65 4 43.48 5.15 59]
Ravindra Jadeja - [30 1.73 4 29.93 4.88 151]
Rishabh Pant - [21 1.7 3 26.1 0 9]
Lokesh Rahul - [27 1.8 3 34.30 0 14]
MS Dhoni - [38 1.75 3 50.72 5.16 341]
Dinesh Karthik - [34 1.7 3 31.03 0 91]
Kuldeep Yadav - [24 1.68 2 12.6 4.93 44]
Mohammed Shami - [29 1.75 2 7.4 5.48 63]
Jasprit Bumrah - [25 1.78 2 3.8 4.51 49]
Bhuvneshwar Kumar - [29 1.75 2 14.6 5.01 105]
Yuzvendra Chahal - [29 1.68 2 7.8 4.89 41]

1. Kumar Sangakara - [41 1.78 3 42 0 404]

By simply looking at the data we can see that Most similar player to Kumar Sangakara is MS Dhoni. The reason for this is that the difference in number of matches played is very large for other players. It is this large that it outweights all the other factors

2. David Warner - [32 1.7 1 46.12 0 115]
By looking at number of matches played only Bhuvneshwar Kumar, Dinesh Karthik & Ravindra Jadeja are somewhat closer to David warner. On calculating nearest neighbour we find that Bhuvneshwar Kumar is actually the most similar player to David Warner.

3) Brett Lee - [42 1.87 2 17.81 4.76 221]
Again because of number of matches outweights all the other factors, so Brett lee is similar to Virat Kohli.

As seen in case of David Warner and Brett Lee, the difference in number of matches played is so large that it outweights all the other factors making the predictions faulty. Thus we need to normalize the distances.

We can use the following weighted distance function:-

$$L = \sqrt{a(\Delta age)^2 + b(\Delta height)^2 + c(\Delta role)^2 + d(\Delta batting\ avg)^2 + e(\Delta bowling\ avg)^2 + f(\Delta \#\ of\ matches)^2}$$

a = 1 | as age has moderate effect on players skill.
b = 0.1 | as height doesn't play much role
c = 10 | A batsman should be similar to a batsman & so on. That's why this thing plays a major role.
d = e = 1 | as they are crutial deciding factors too.
f = 0.1 | Experience does play a good role in finding similarity amongst players but it shouldn't be the deciding factor in finding similarity.

Q3)

1) Let $w = [a \quad b \quad c]^T$.

$w^T x_1 < 0 \Rightarrow a + 2b + 3c < 0 \longrightarrow (1)$

$w^T x_2 > 0 \Rightarrow \cancel{2a+4b+4c>0} \quad 3b + 4c > 0 \longrightarrow (2)$

$w^T x_3 > 0 \Rightarrow 2a + 4b + 4c > 0 \longrightarrow (3)$

$(1) \times -2$

$\Rightarrow -2a - 4b - 6c > 0 \longrightarrow (4)$

$(3) + (4)$

$-2c > 0$

$\Rightarrow c < 0$.

$\Rightarrow 3b < -4c$

$\Rightarrow b < -\frac{4}{3}c$.

$a < -2b - 3c$

$\therefore$ ~~te~~ Since we need only one hyperplane, let $c = -1$.

$\Rightarrow b < 4/3 \Rightarrow b = 1$

$a < -2 + 3$

$\Rightarrow a < 1$

$\Rightarrow a = 0$

$\therefore w = [0 \quad 1 \quad -1]^T$.

2) Let $x_1 = [\alpha_1 \ \alpha_2 \ \alpha_3 \ \alpha_4]$ $\quad \alpha \in R^4$

$x_2 = [2\alpha_1 \ 2\alpha_2 \ 2\alpha_3 \ 2\alpha_4]$

$x_3 = [3\alpha_1 \ 3\alpha_2 \ 3\alpha_3 \ 3\alpha_4]$

$x_4 = [4\alpha_1 \ 4\alpha_2 \ 4\alpha_3 \ 4\alpha_4]$.

Now let us take w.

$\theta_i = \therefore \ w^T x_i \quad i \in \{\cancel{0}, 1, 2, 3, 4\}$.

~~= i w^T x_1 ie~~

$\theta_i = i w^T x_1 \quad i \in \{1, 2, 3, 4\}$

Now ~~i~~

Now Let $\theta_1 = w^T x_1$.

$\therefore \theta_i = i \, \theta_1$

Since ~~θ~~ i is positive integer. θ_i and θ_1 would have same sign and thus lies on the same side of the hyperplane. Thus for all values of ~~θ_i~~ i, the point will lie on same side of hyperplane and thus there exits no w that can keep ~~x~~ x_1 & x_2 on one side and x_3 & x_4 on other side.