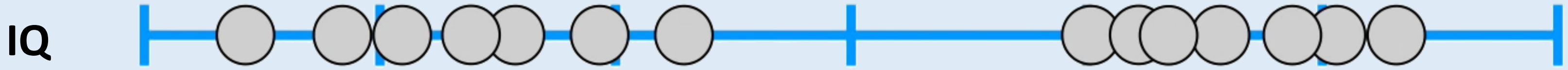
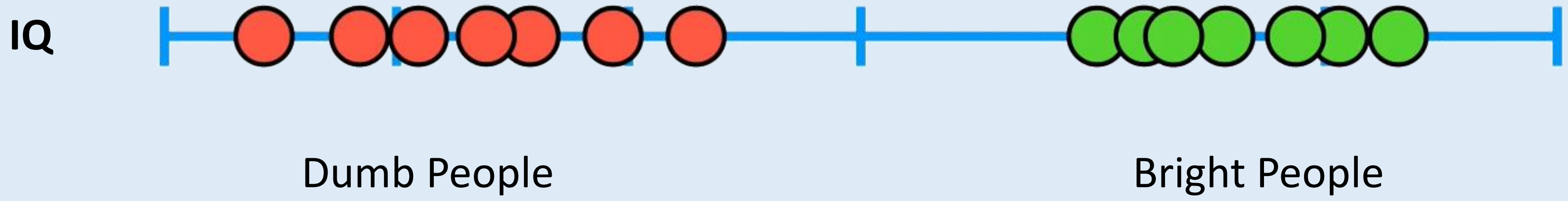
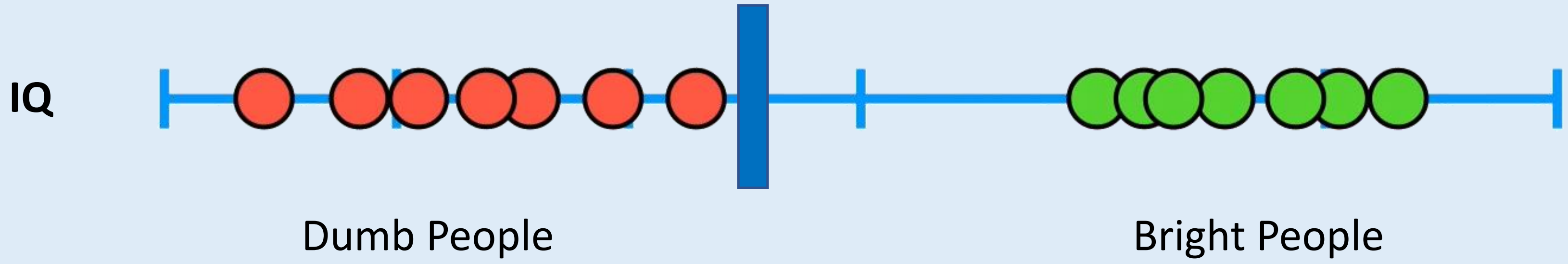
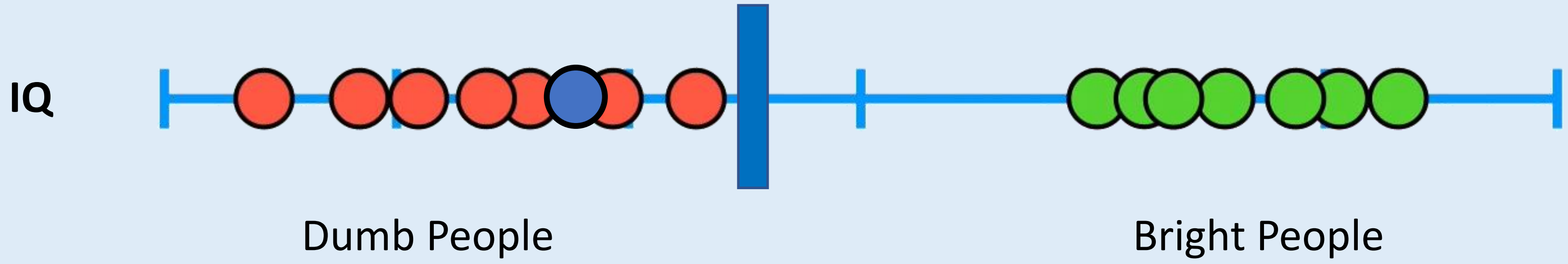
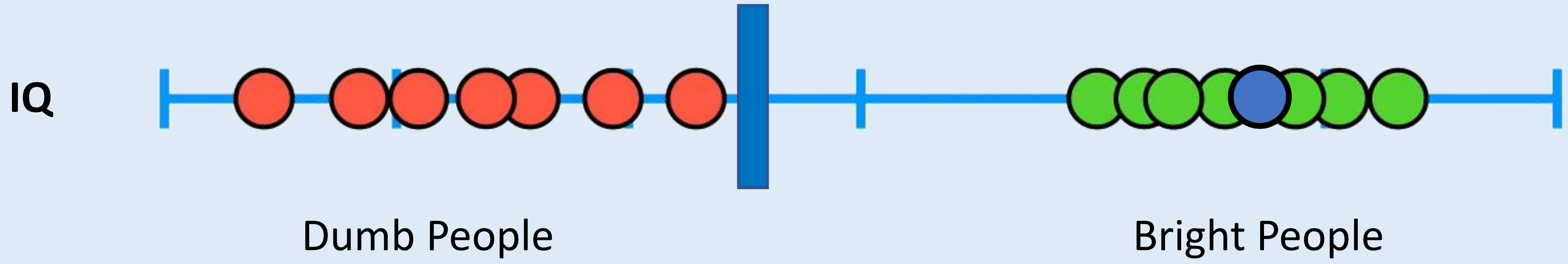
Support Vector Machine

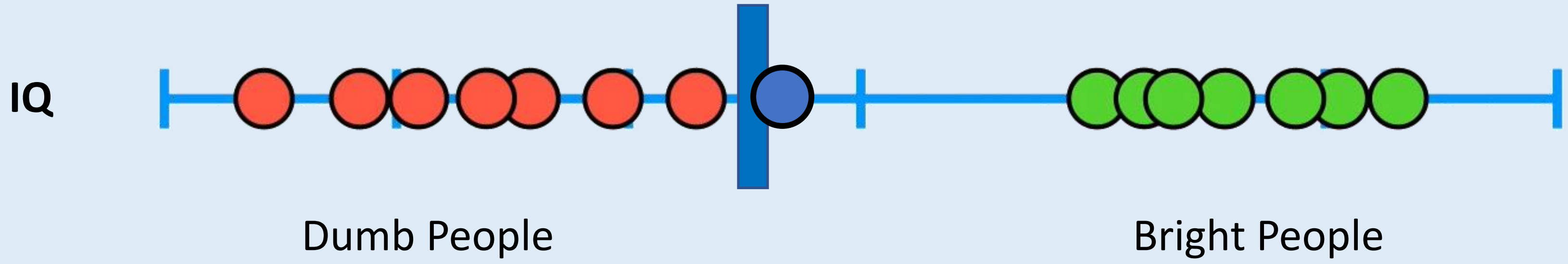

IQ

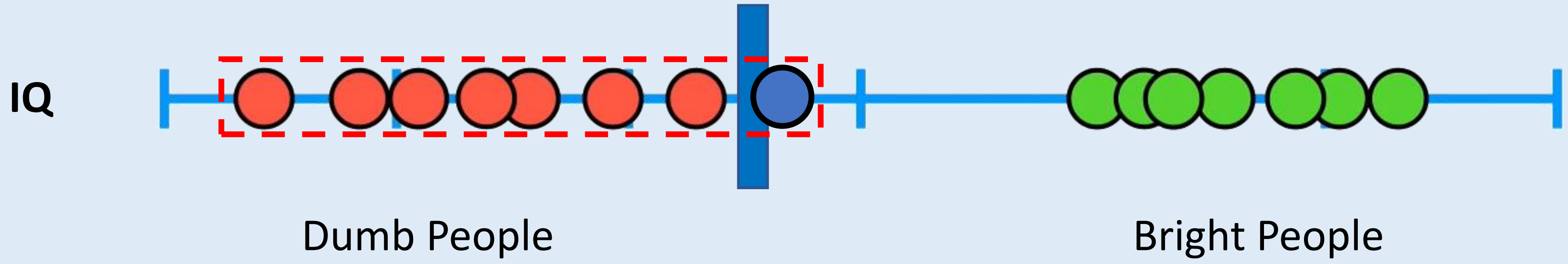

IQ
Dumb People
Bright People

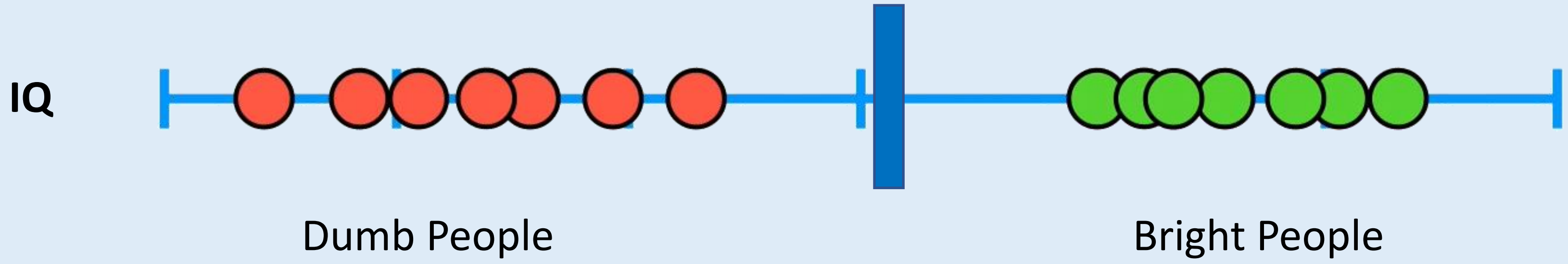

IQ
Dumb People
Bright People

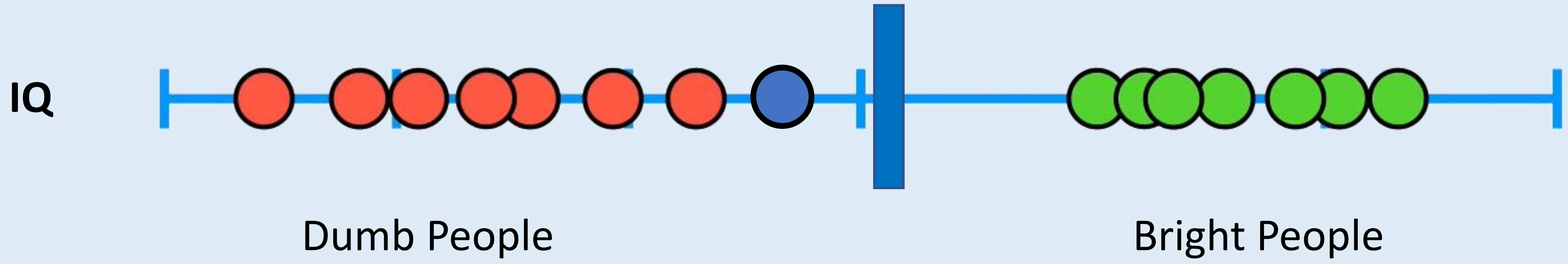

IQ
Dumb People
Bright People

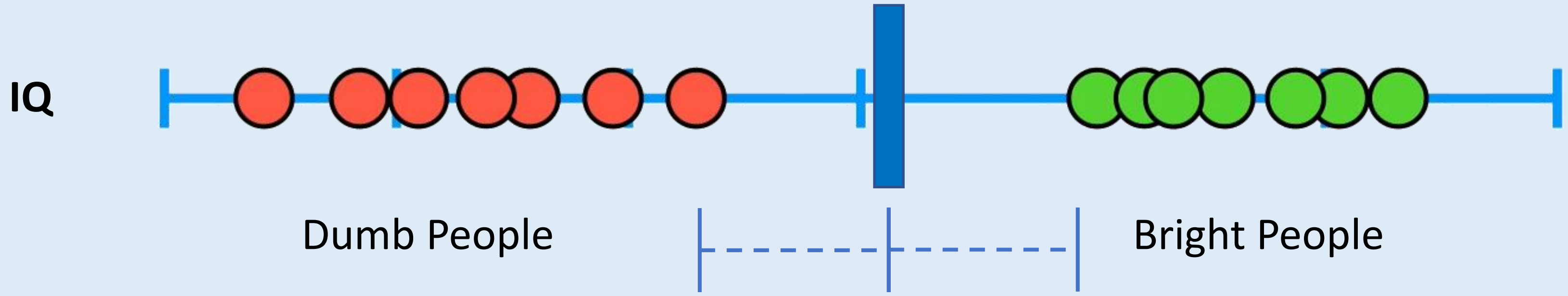

IQ
Dumb People
Bright People


IQ
Dumb People
Bright People

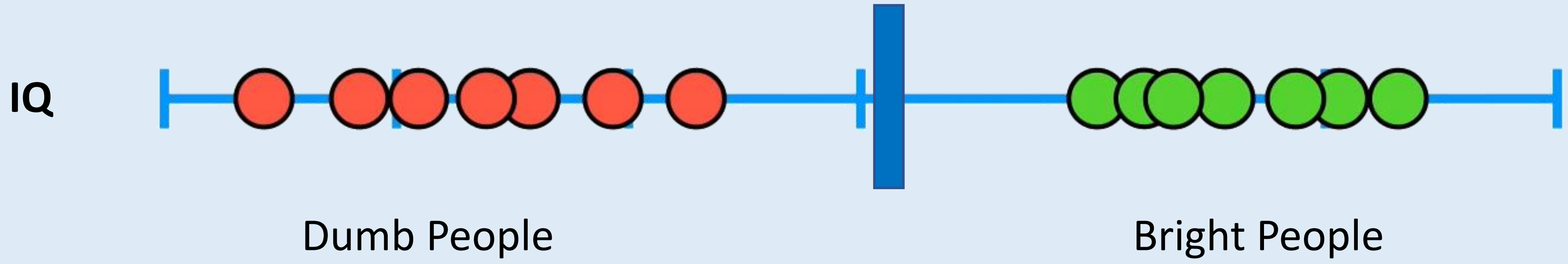

IQ
Dumb People
Bright People


IQ
Dumb People
Bright People

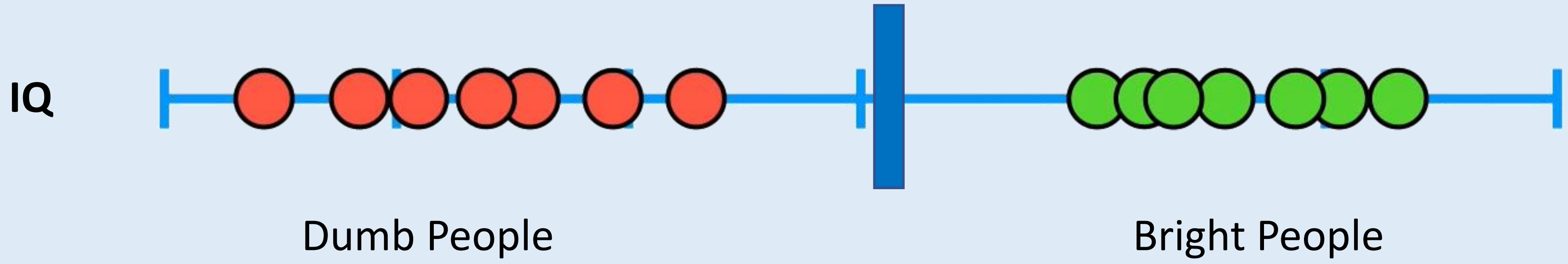

IQ
Dumb People
Bright People



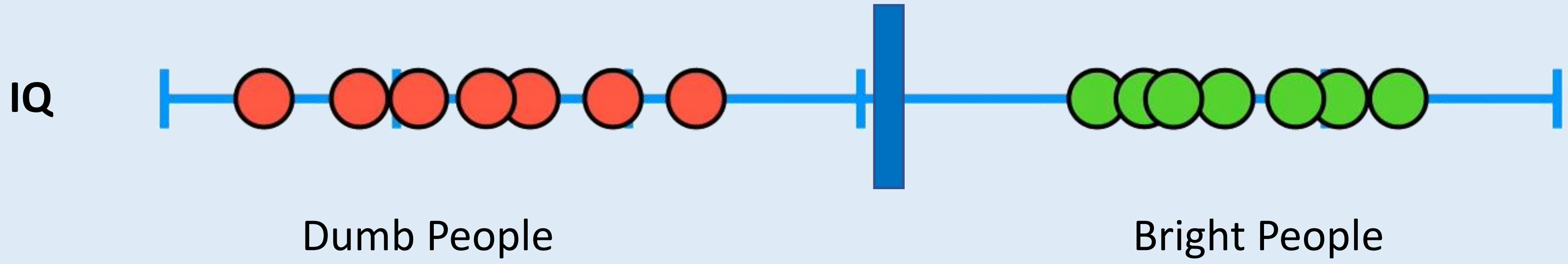
Shortest Distance between threshold and the observation is called margin



Shortest Distance between threshold and the observation is called margin

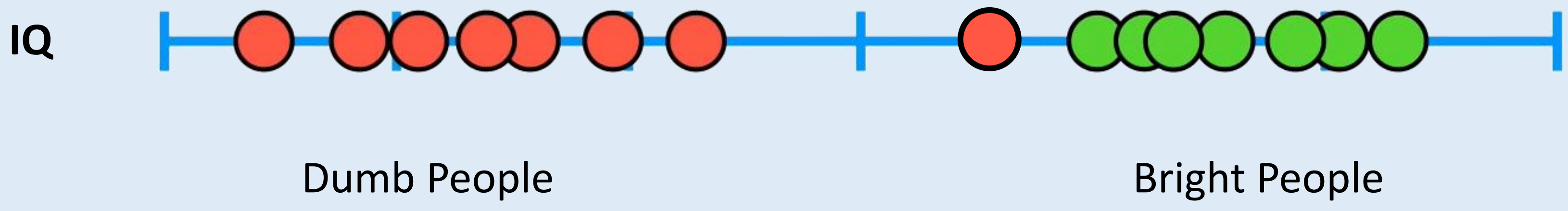


If we move the threshold to either side i.e. towards observation on right or observation on left, the margin would reduce or it will be called a smaller margin



If the threshold stays at the position at which it is right now, the distance between both observation is maximum,
The arrangement will be called as maximum margin classifier

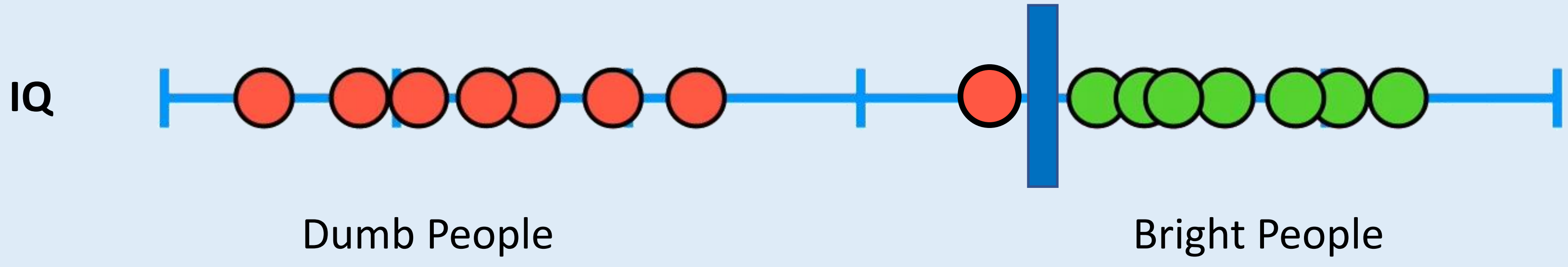
Is maximum margin classifier good enough?



If the threshold stays at the position at which it is right now, the distance between both observation is maximum,
The arrangement will be called as maximum margin classifier

Is maximum margin classifier good enough?

What if the training data looks like this?

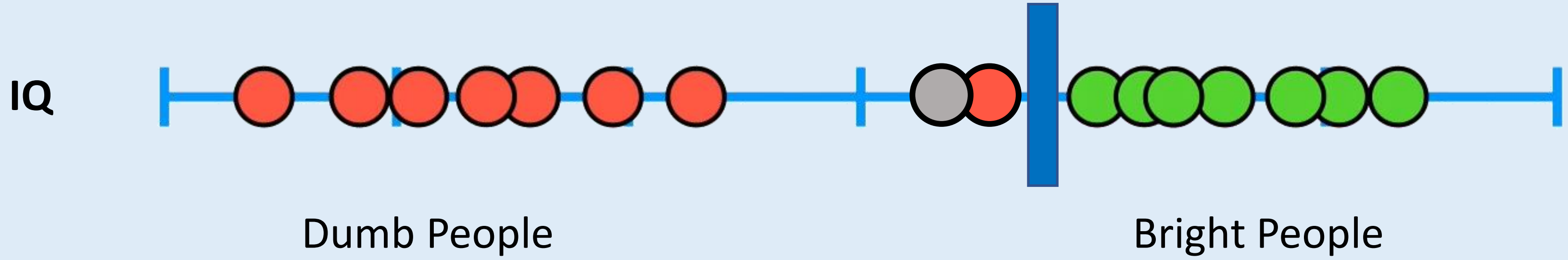


If the threshold stays at the position at which it is right now, the distance between both observation is maximum, The arrangement will be called as maximum margin classifier

Is maximum margin classifier good enough?

What if the training data looks like this?

Maximum margin classifier will be close to Bright People… and far from general observation of dumb people…



Let’s say the new observation looks like this…

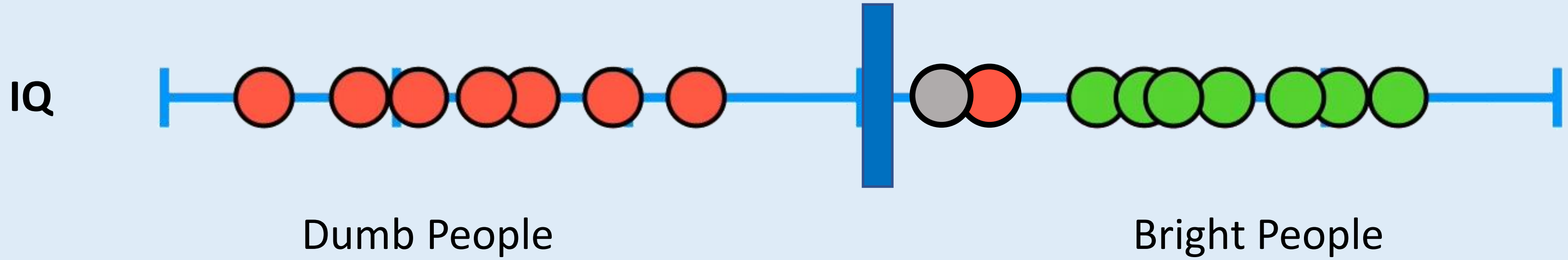
The observation is an outlier of dumb class….

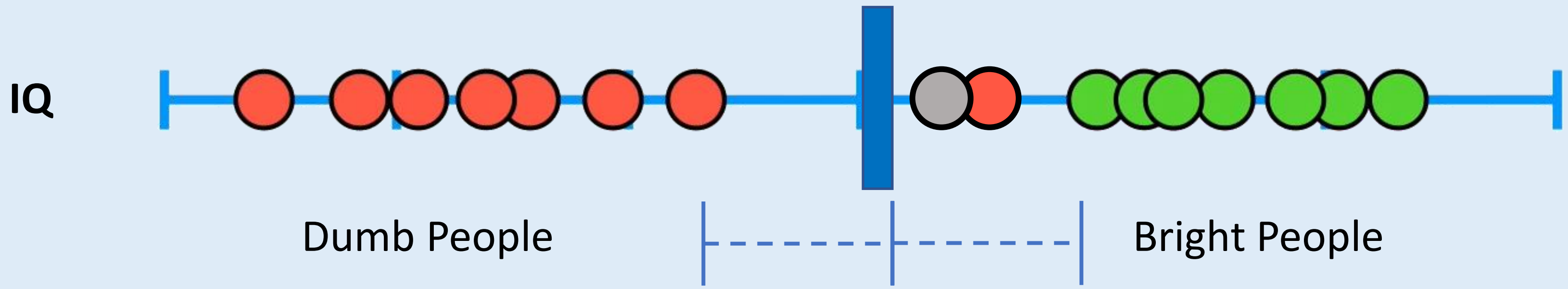
So now even if the new observation belongs to bright people, it will be classified as dumb person

Therefore the Maximum Margin Classifier are sensitive to outlier…

How to correct this mistake?

To correct this mistake, we should allow miss classification.. With the previous arrangement of threshold

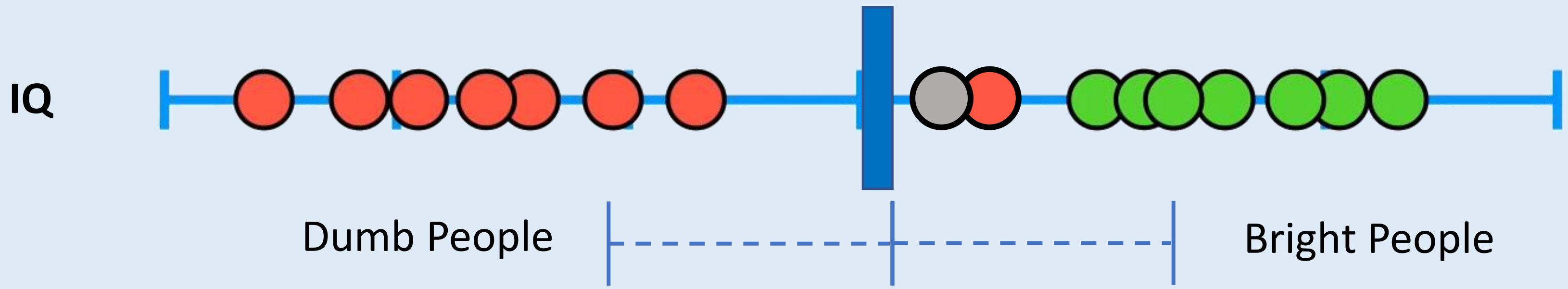




Now by rearranging the threshold we have lowered the variance in the model and increased the biased

The distance which allows the misclassification is called as soft margin

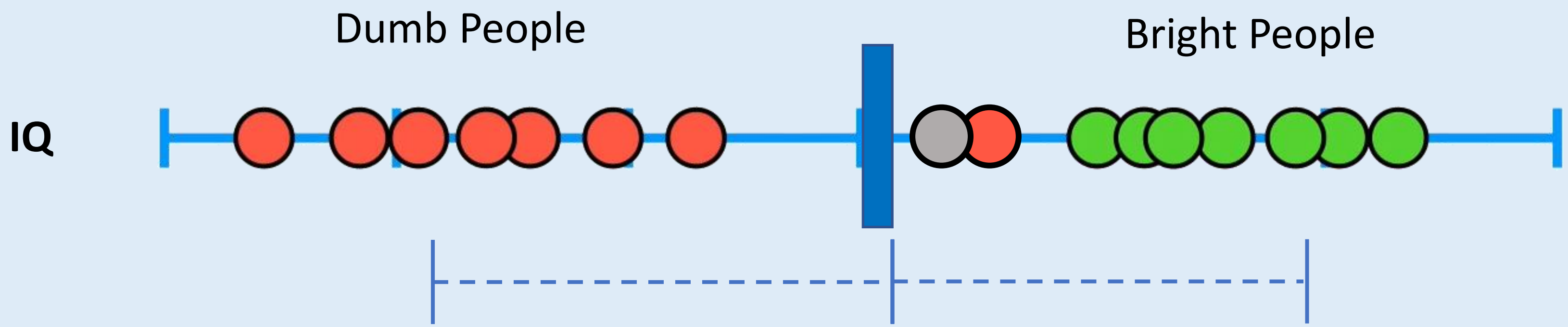
Now there could be several soft margins which allows misclassifications and determined by other observation pair instead of the extreme last of dumb people and first of bright people
So the observation that could be considered for forming new soft margin could be second last of dumb people and second of Bright people
Or
Third last of dumb people and third of bright people .. So on and so forth..
So which has to be chosen as best one?
For that we can use cross validation and choose the best soft margin…



Now by rearranging the threshold we have lowered the variance in the model and increased the biased

The distance which allows the misclassification is called as soft margin

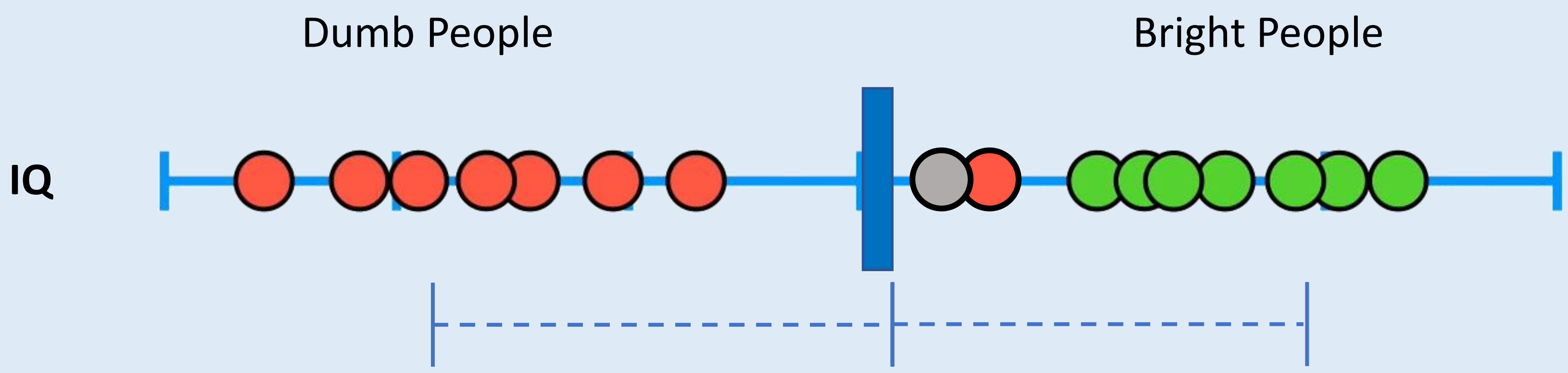
Now there could be several soft margins which allows misclassifications and determined by other observation pair instead of the extreme last of dumb people and first of bright people
So the observation that could be considered for forming new soft margin could be second last of dumb people and second of Bright people
Or
Third last of dumb people and third of bright people .. So on and so forth..
So which has to be chosen as best one?
For that we can use cross validation and choose the best soft margin…



Now by rearranging the threshold we have lowered the variance in the model and increased the biased

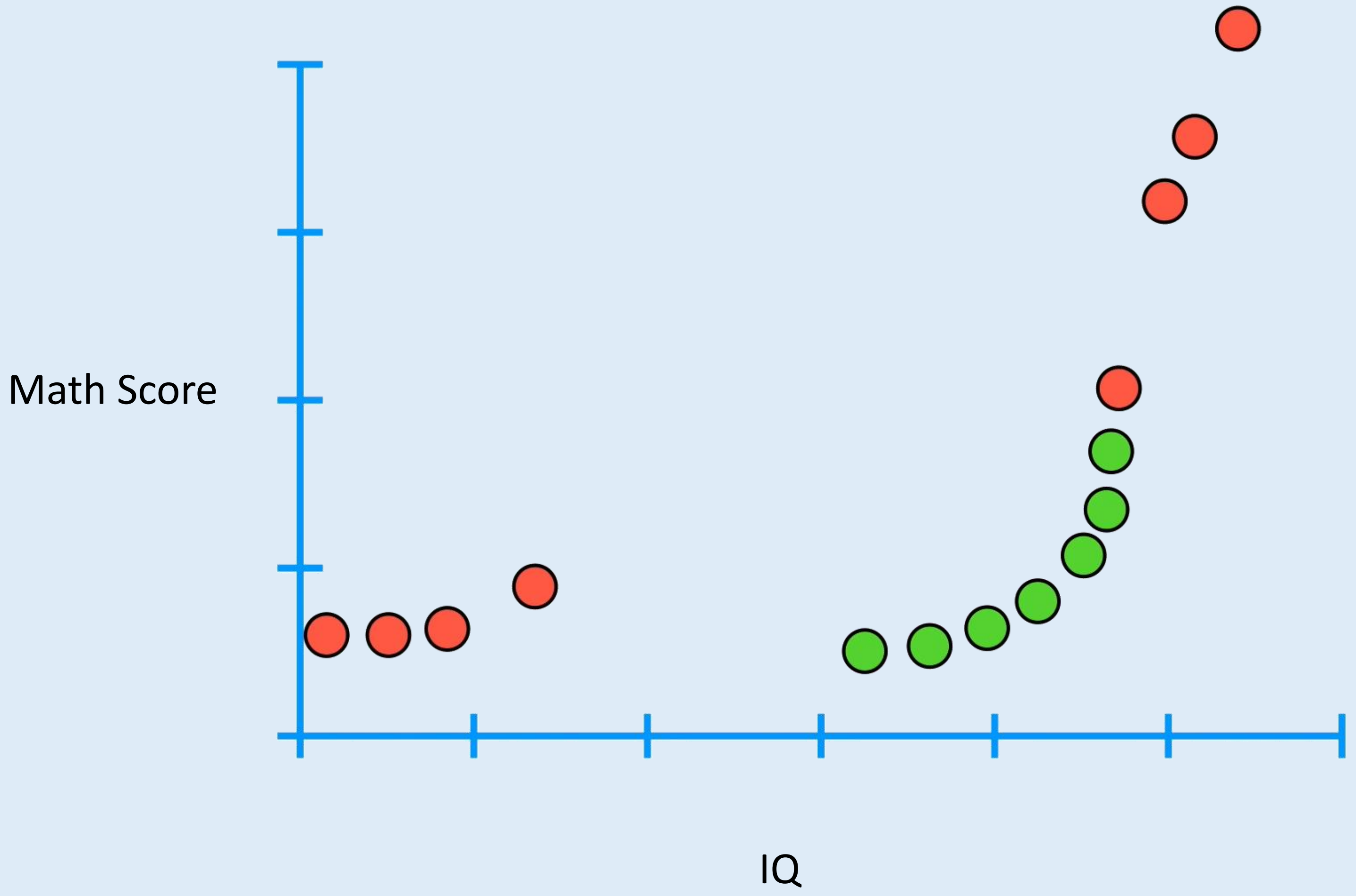
The distance which allows the misclassification is called as soft margin

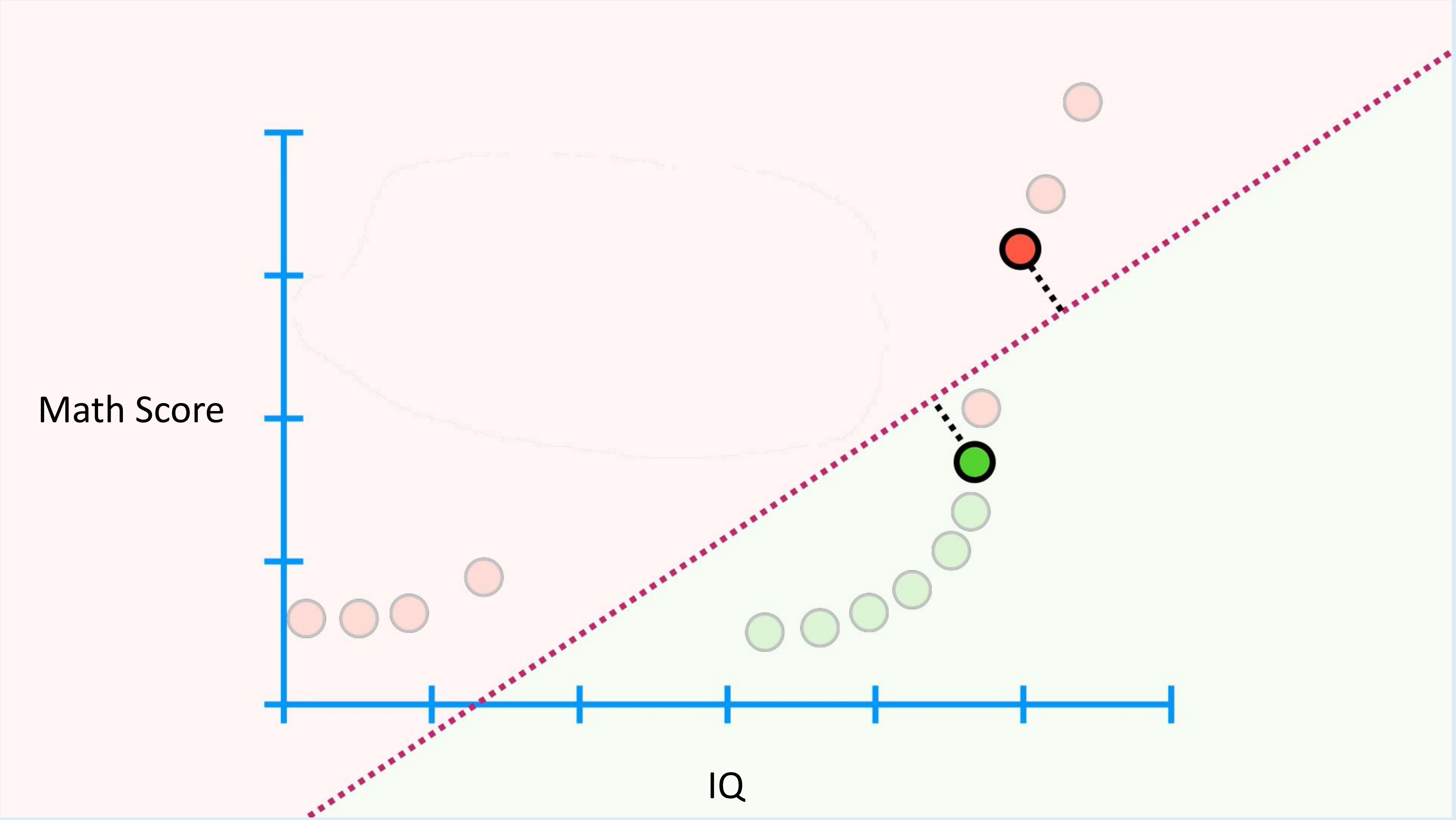
Now there could be several soft margins which allows misclassifications and determined by other observation pair instead of the extreme last of dumb people and first of bright people
So the observation that could be considered for forming new soft margin could be second last of dumb people and second of Bright people
Or
Third last of dumb people and third of bright people .. So on and so forth..
So which has to be chosen as best one?
For that we can use cross validation and choose the best soft margin…



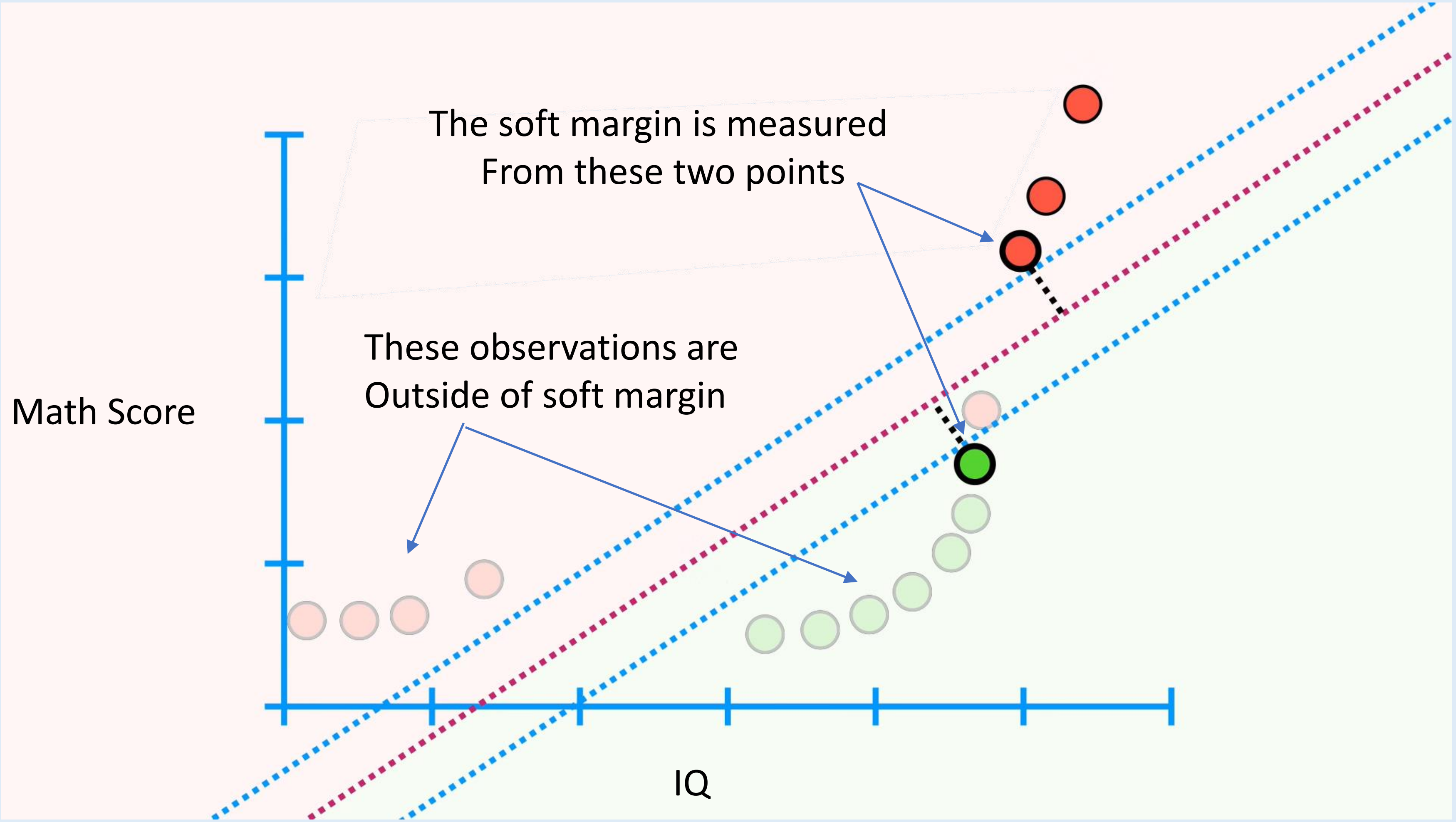
Now we can decide the threshold using this soft margin
And the classifier made using such soft margin is called as soft margin classifier or
Support vector classifier

The names are derived form the observation on the edge and within the soft margin or the support vectors

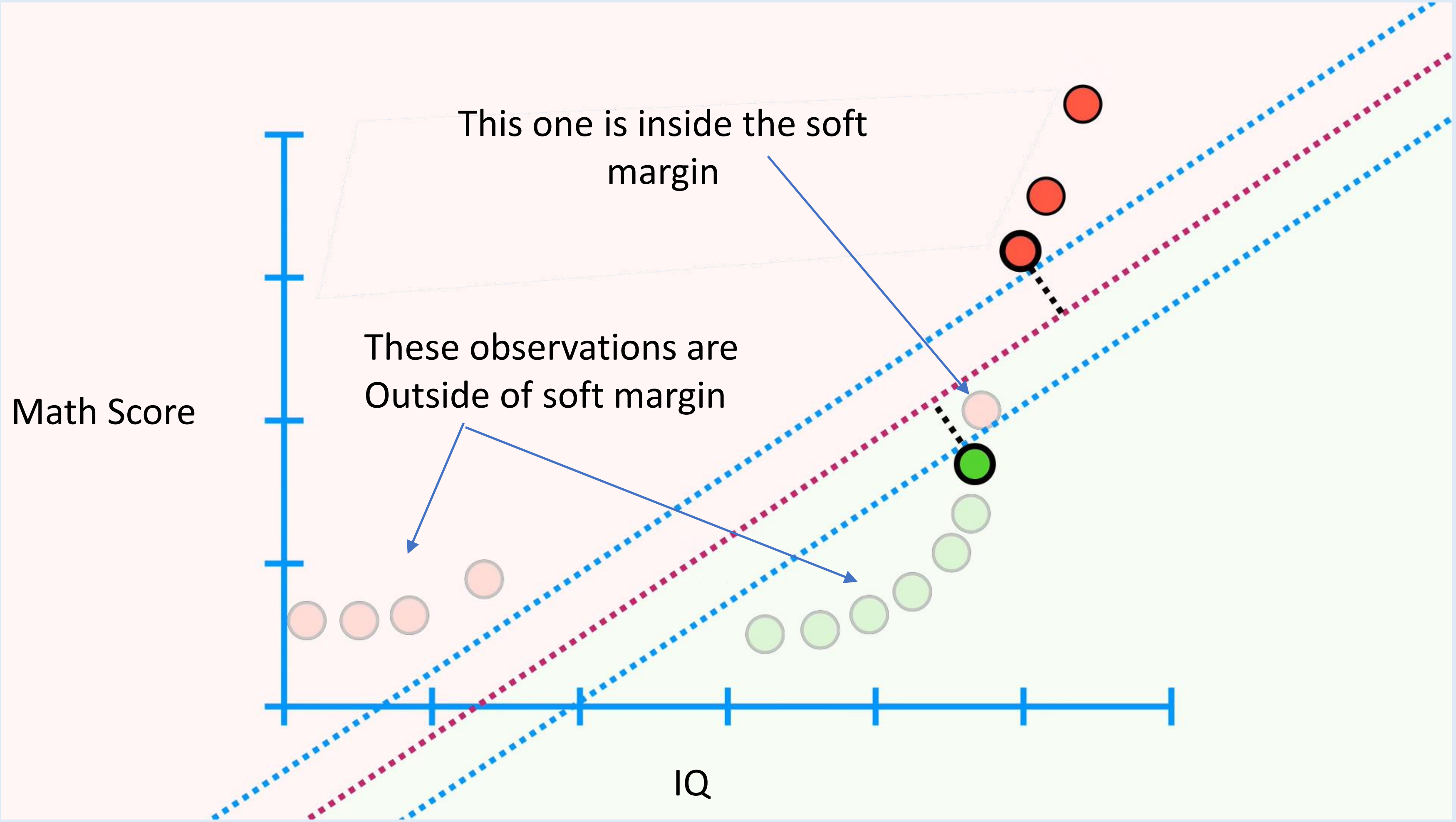

Math Score
IQ



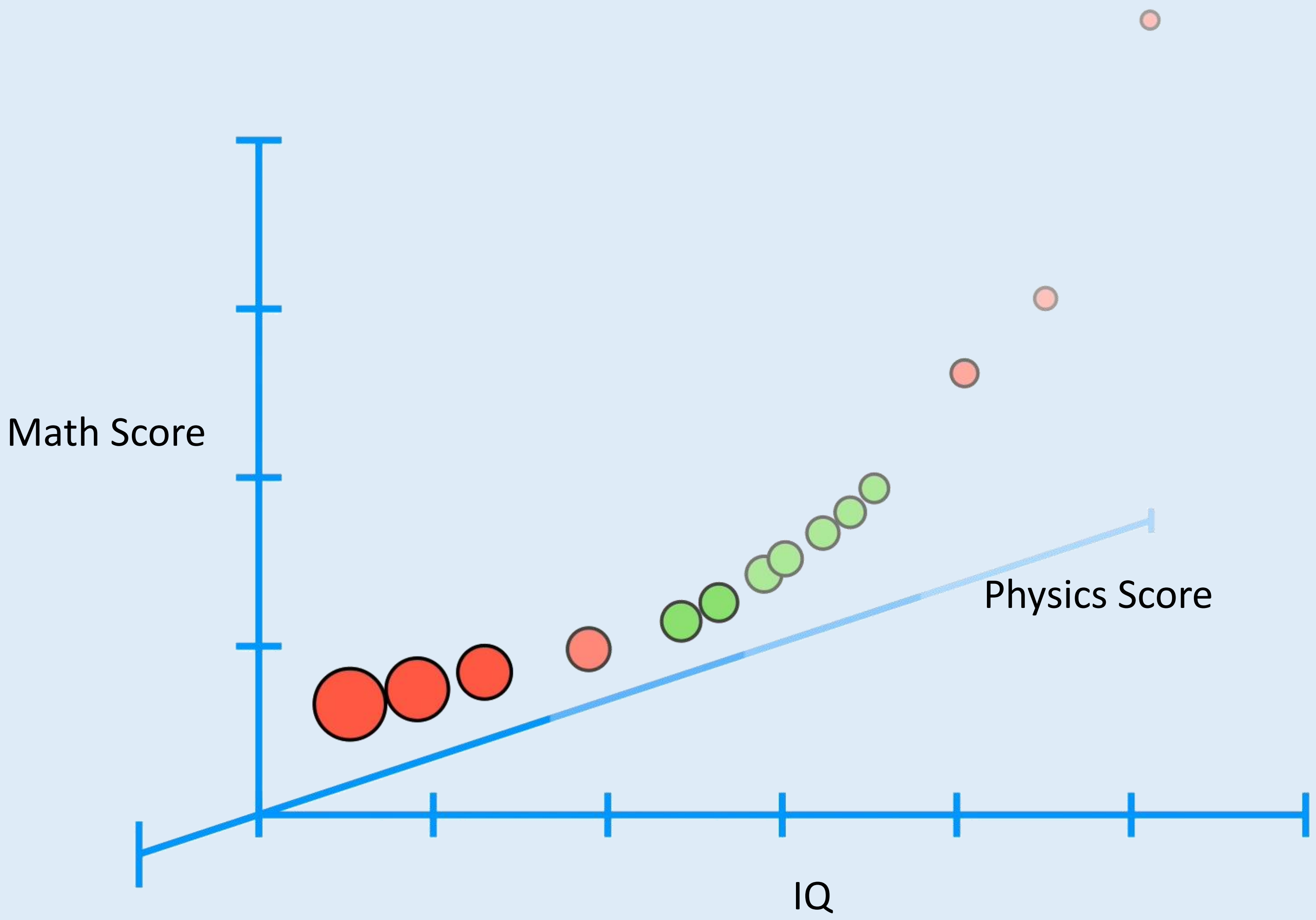
For two dimensional data the support vector classifier will be a line



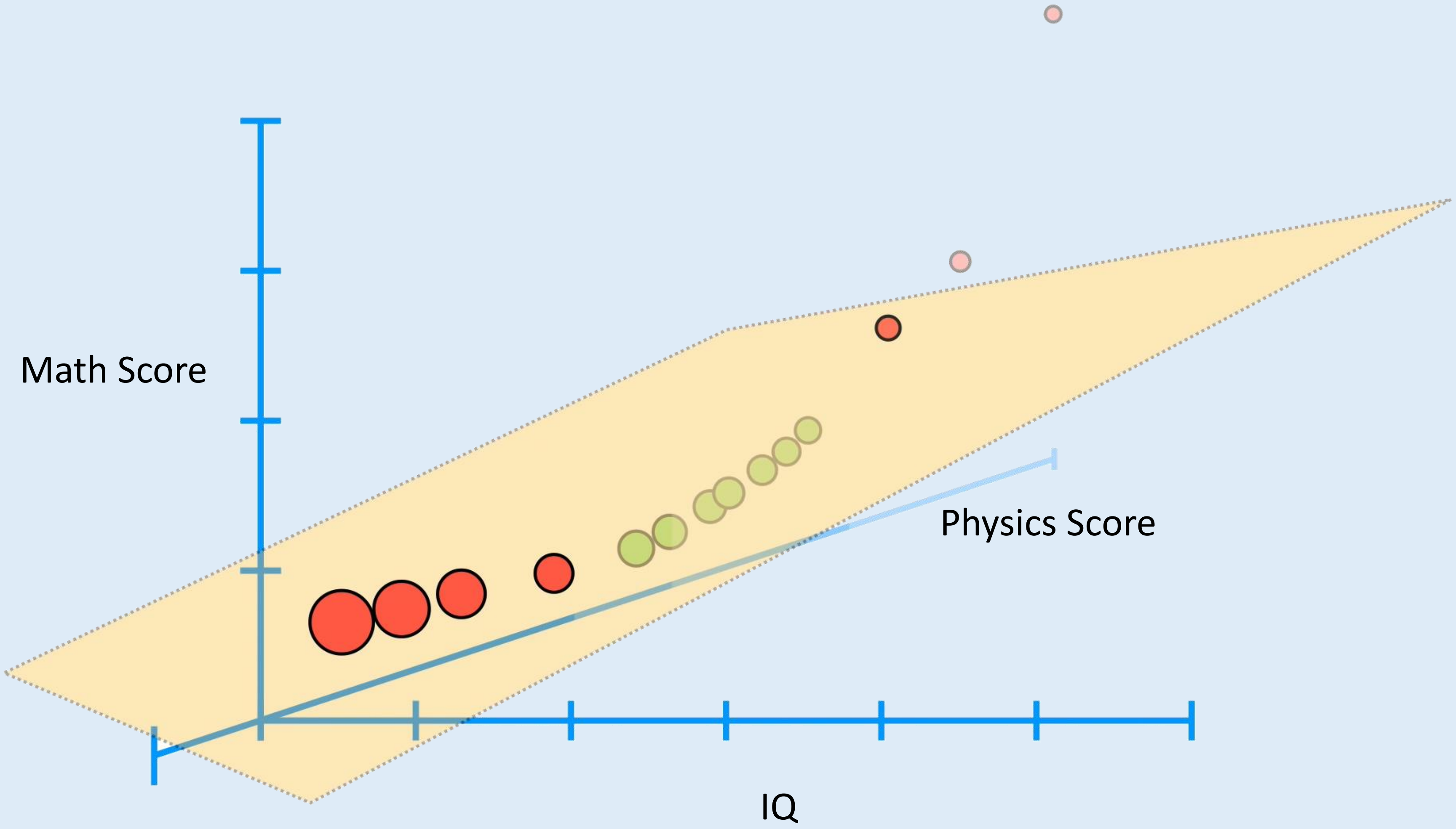
The blue lines gives us the idea about where the other point are as compered to the soft margin



The blue lines gives us the idea about where the other point are as compered to the soft margin



Now the data is three dimensional

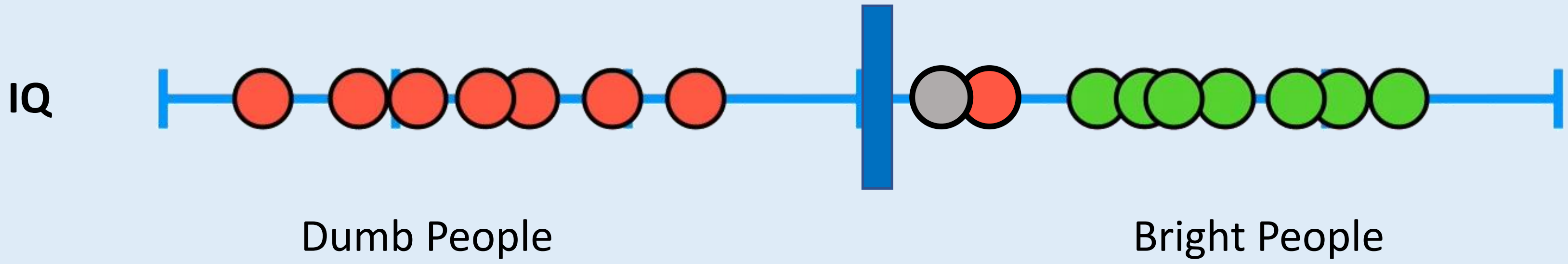


For three dimensional data, the support vector classifier is a plane, Which will help us to classify the bright and dumb people

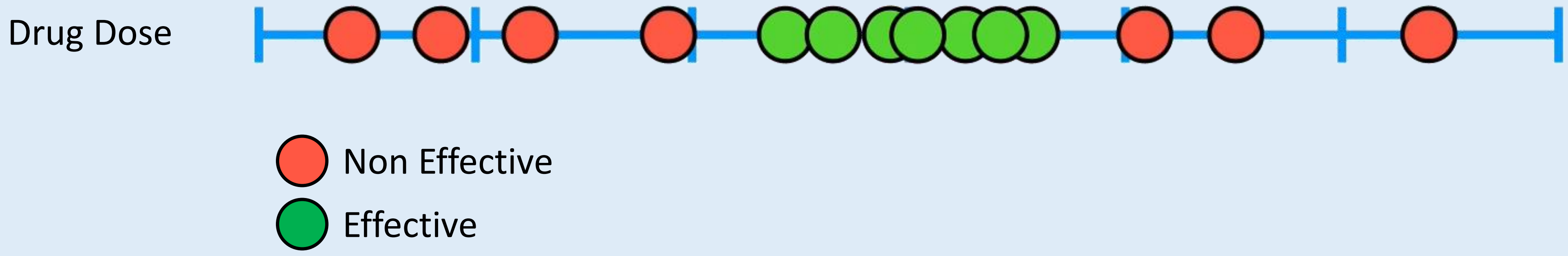
To summarise:

- If the data is one dimensional >> The support vector classifier is a point
- If the data is two dimensional >> The support vector classifier is a one dimensional line
- If the data is three dimensional >> The support vector classifier is a two dimensional plane
- If the data is 4 dimensional or higher ?
- If the data is 4 dimensional or higher >> The support vector classifier is a hyperplane

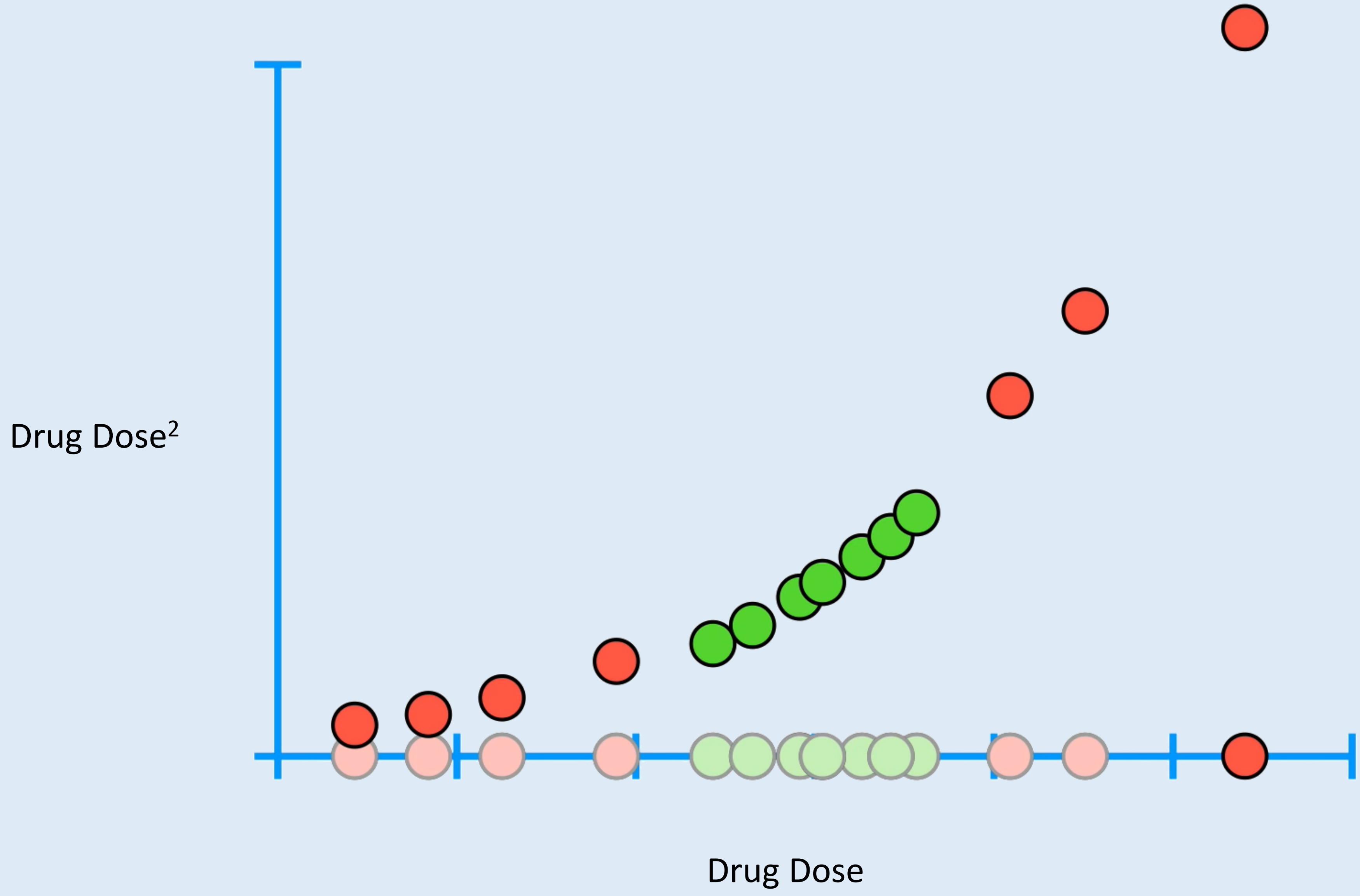
- The support vector classifier worked well with such simple data
- It also handled the outliers well and allowed the misclassification
- Since the data was quite straight forward the Support Vector Classifier did not have any trouble working



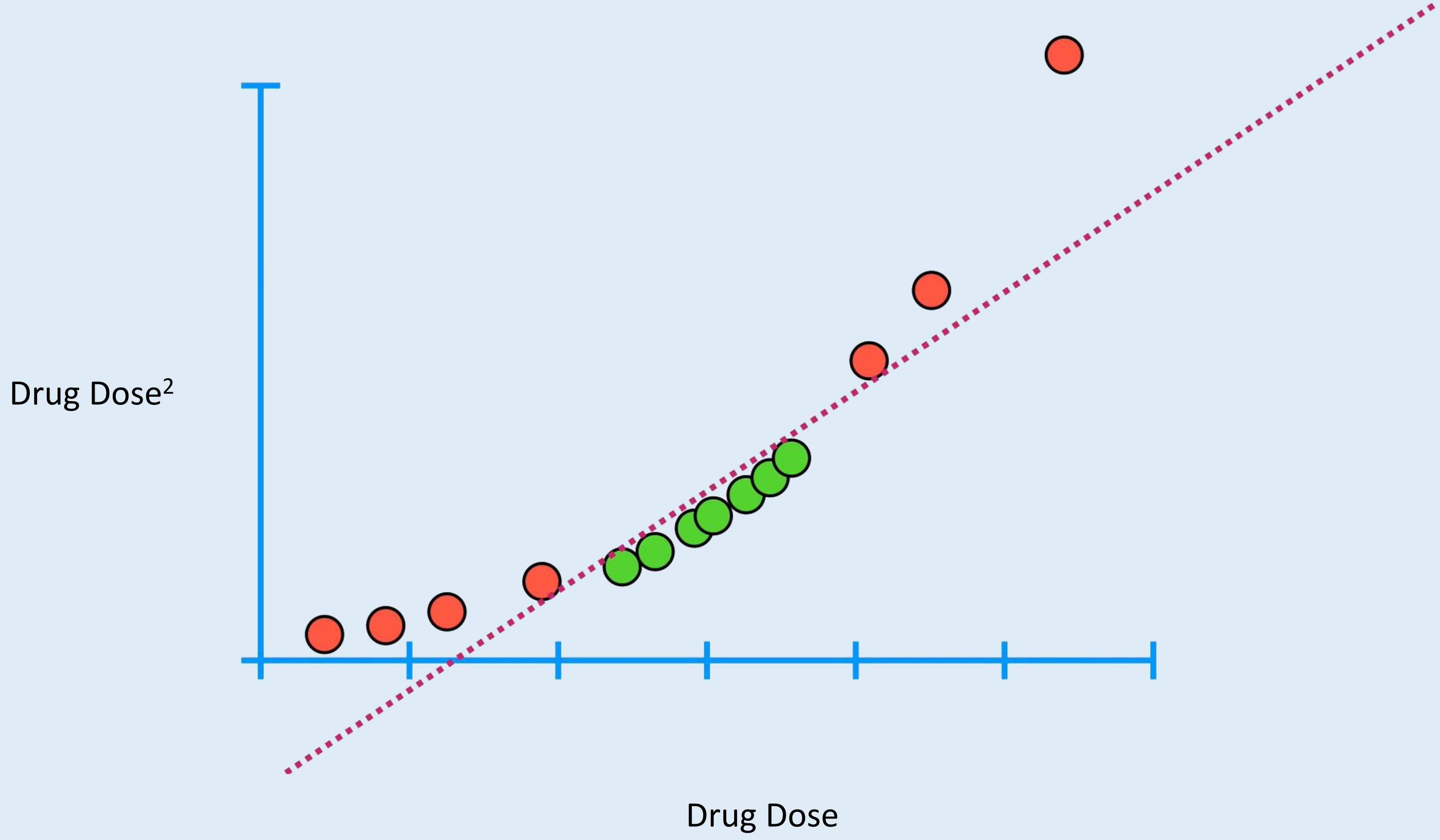
- What if the data was like this?
- Now, for such data no matter where we put the classifier, we cannot classify the data effectively
- So the support vector classifier worked well with the data which was easy to classify and did not have overlapping points
- Then how to classify such points?
- Here the Support Vector Machine helps us…



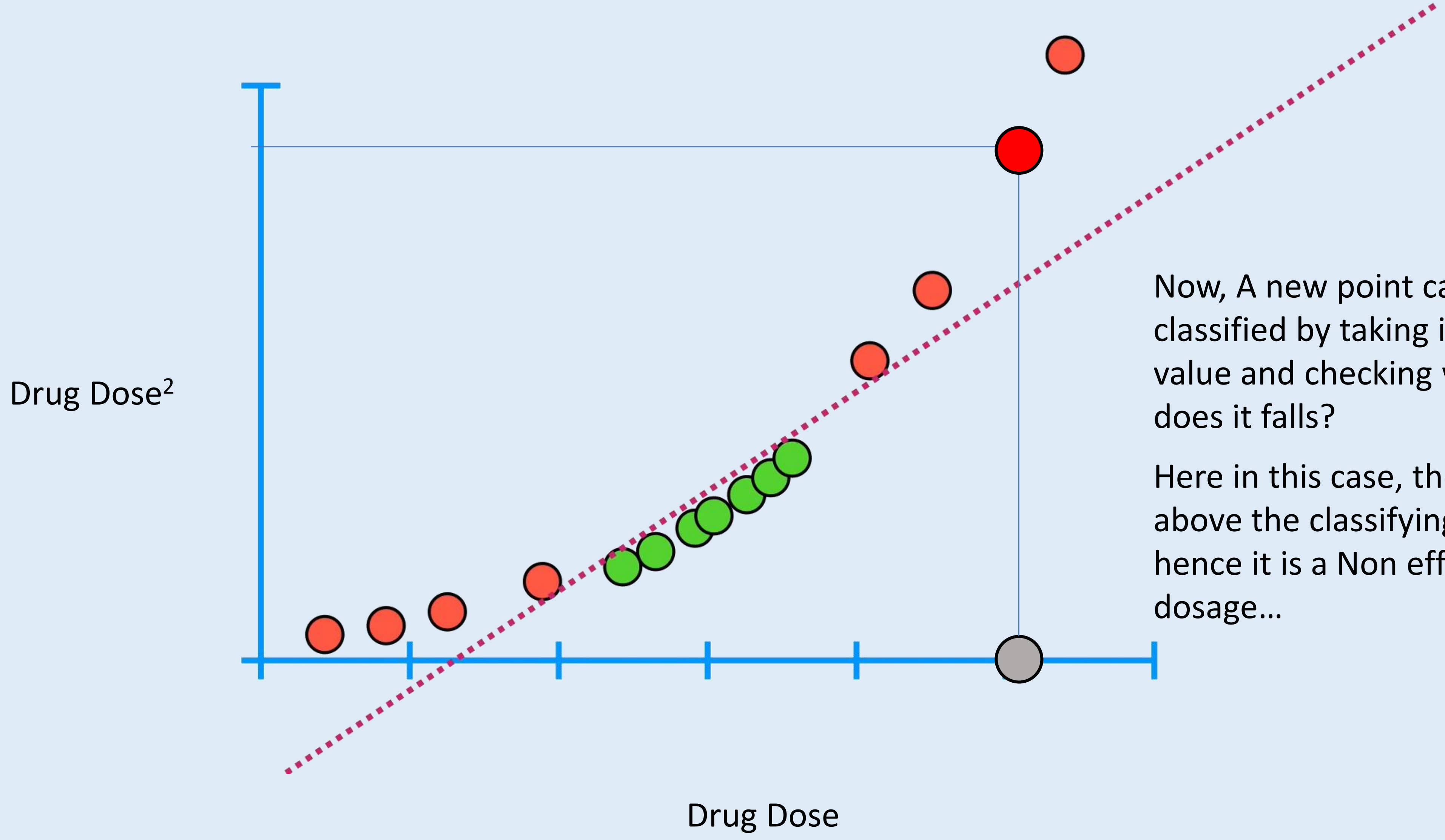

Non Effective
Effective
y
Drug Dose²
Drug Dose



Drug Dose^2
Drug Dose



Drug Dose^2
Drug Dose

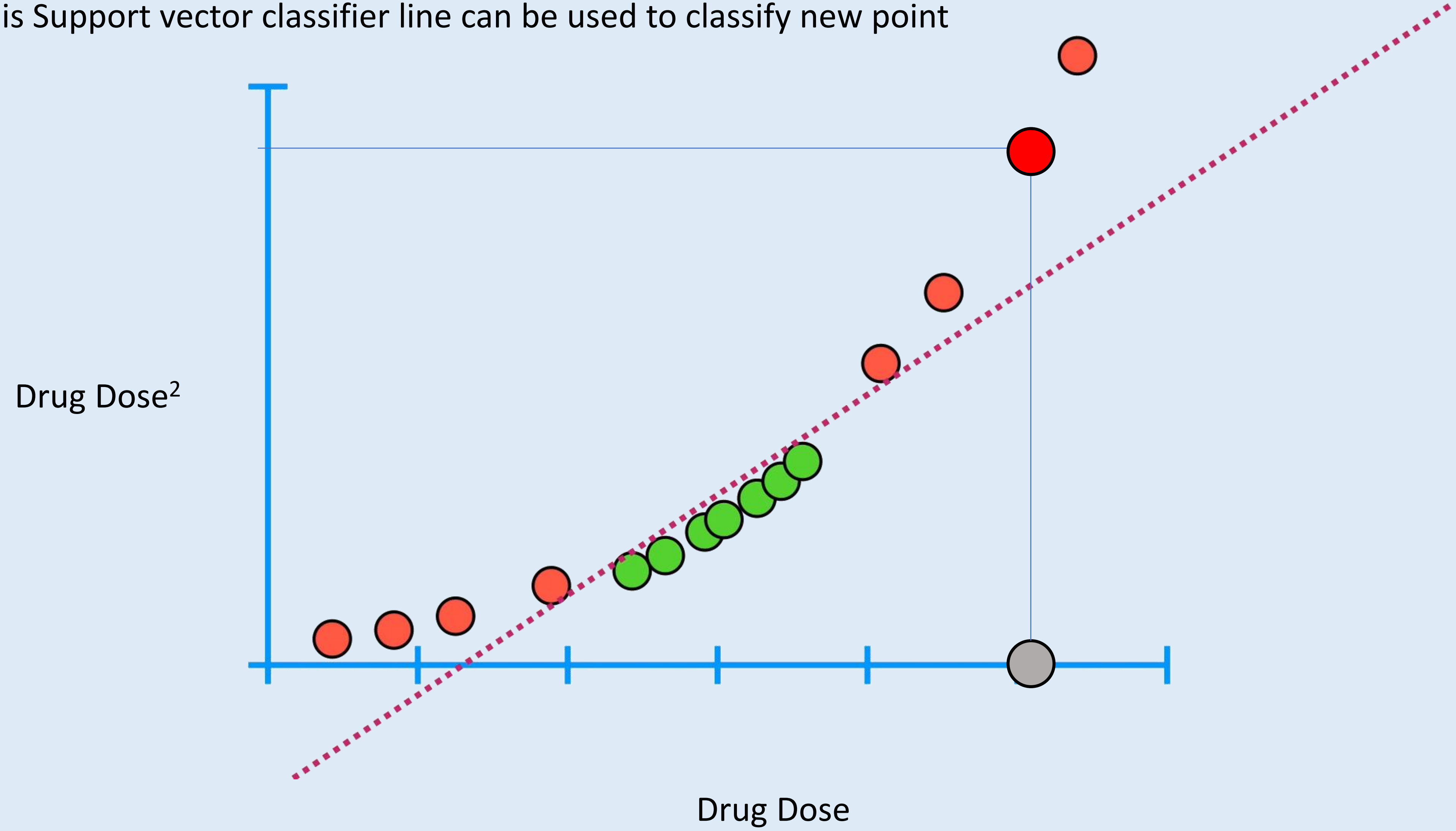


Now, A new point can be classified by taking its square value and checking where does it falls?

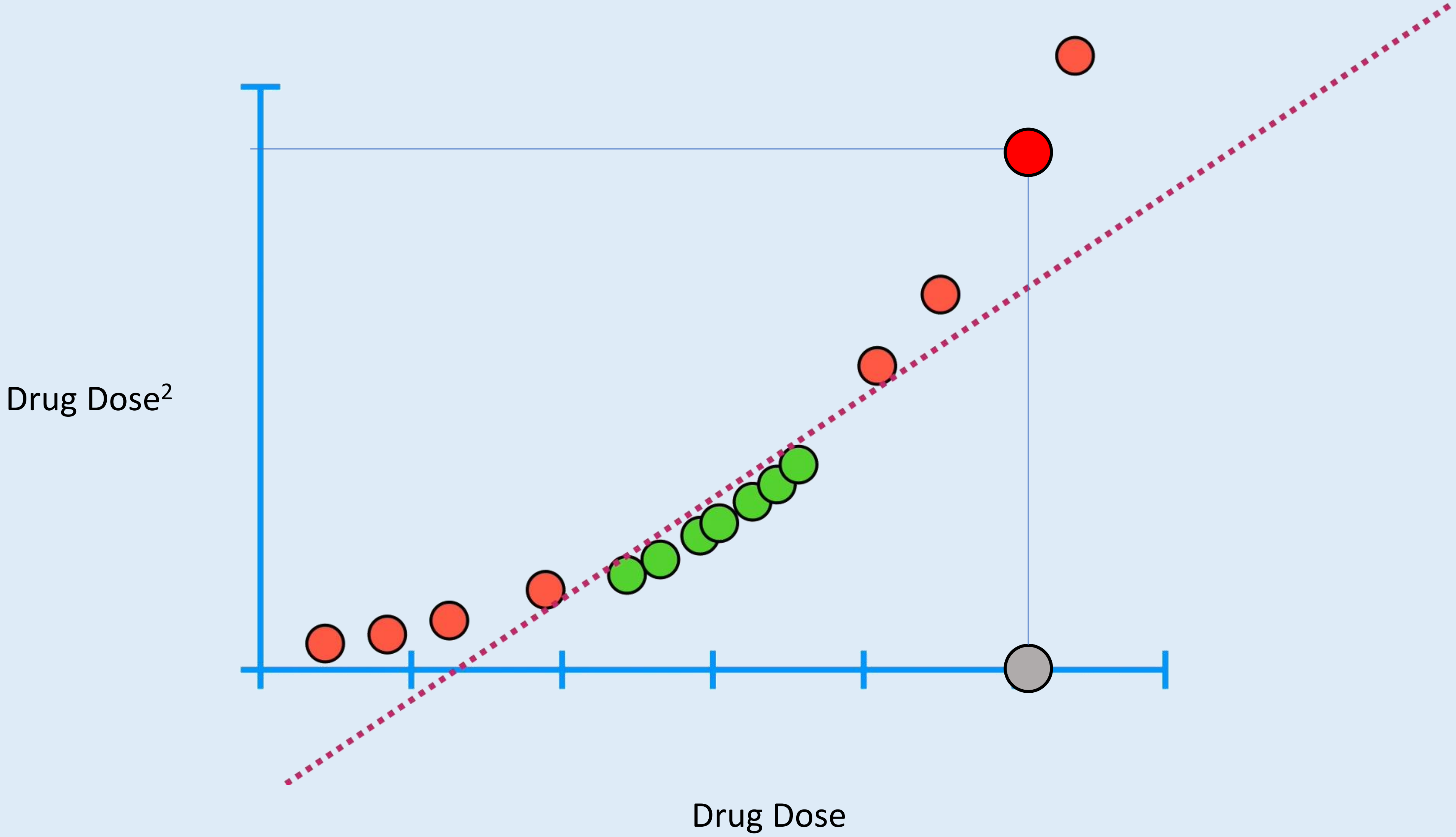
Here in this case, the point is above the classifying line hence it is a Non effective dosage…

To summarise:

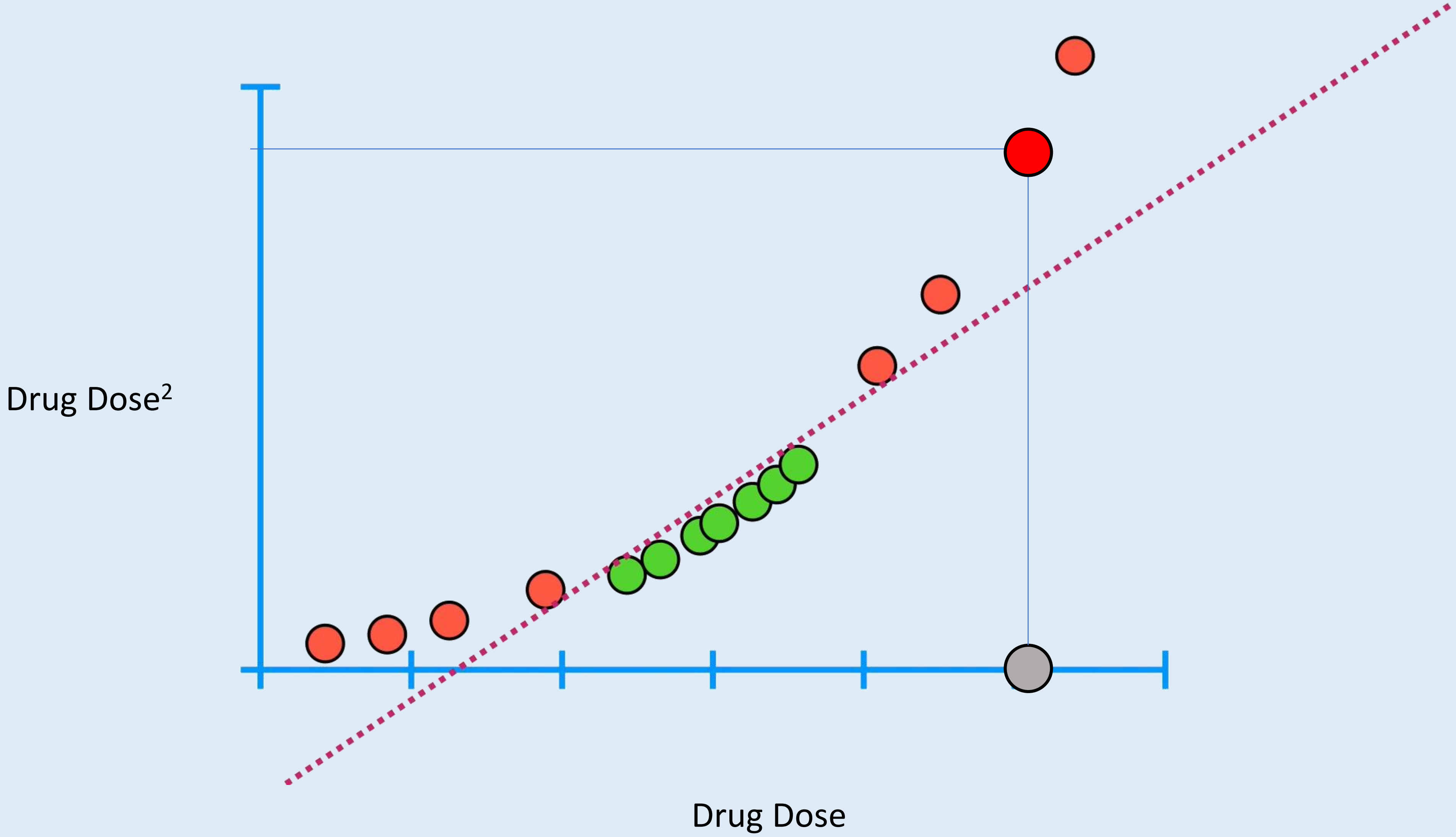
- Started with a data with lower dimension (Here in this example it is one dimensional data)
- The data was transferred to the higher dimension (Data was transferred to 2D)
- Support vector classifier was found that separates the data in two group (Here, a line was created)
- This Support vector classifier line can be used to classify new point



- Now you must be wandering why did we transform data by square only and why not cube
- Some might be wandering why we did not transform it by taking square or cube root
- Now, how to decide which data transform is to be used?
- Support Vector Machines use Kernel Function to identify which support vector classifier to use in higher dimension

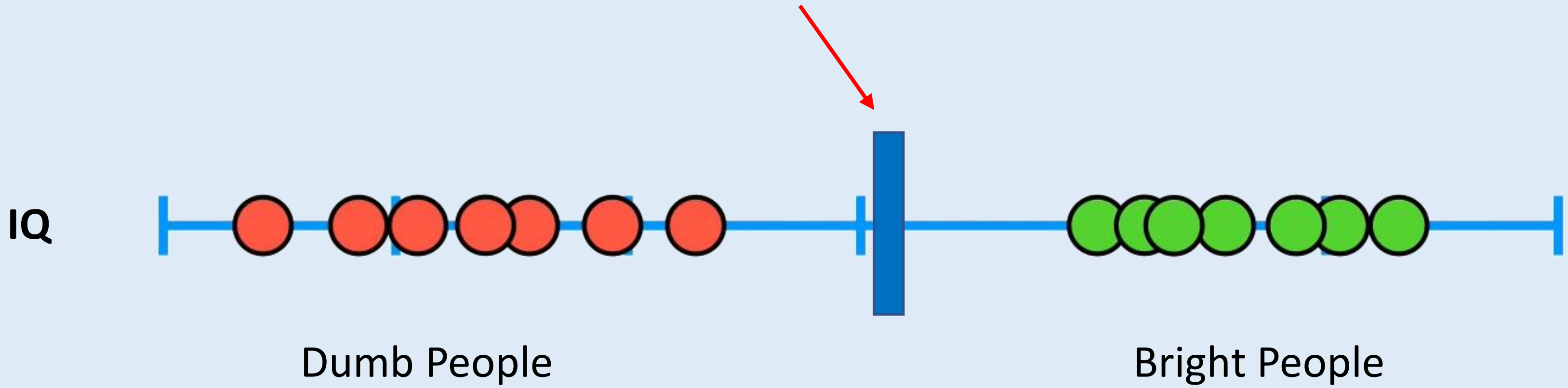


Here, Polynomial Kernel function is used to find the support vector classifier
The Polynomial Kernel Function has a Parameter d
The parameter d decides the degree of polynomial of Kernel Function

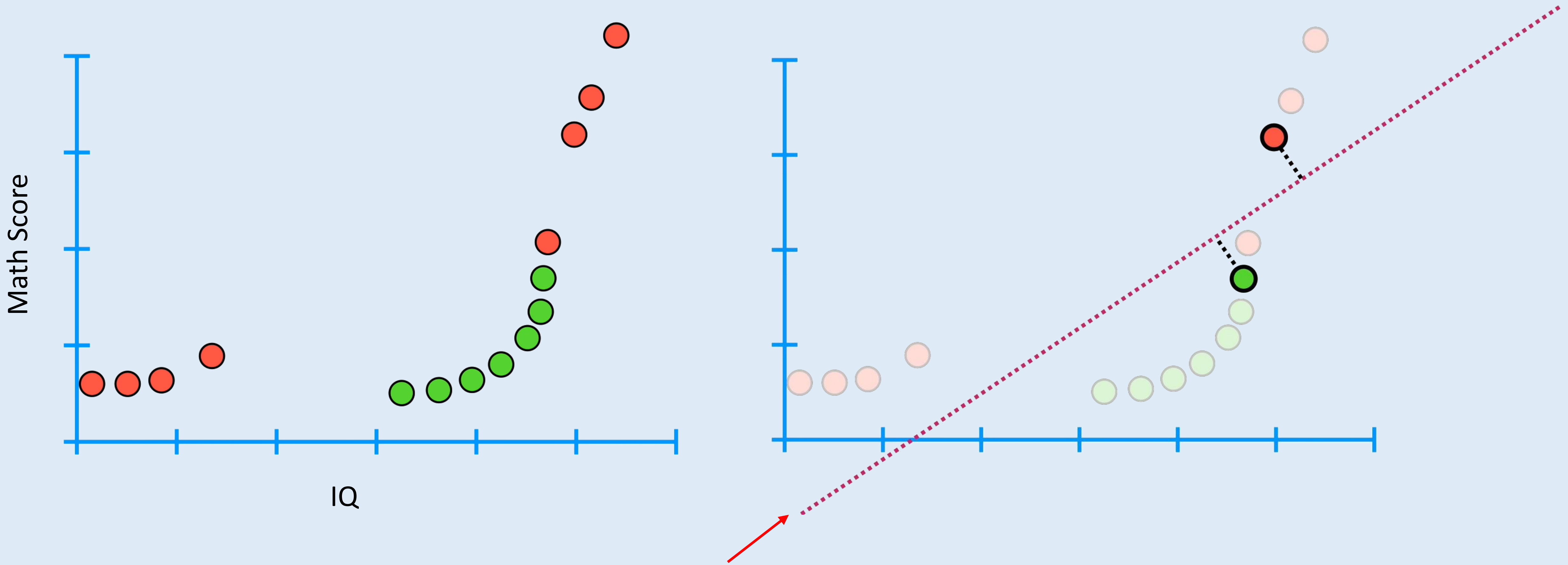


When d = 1, the Polynomial Kernel Function decides the relationship between each pair of datapoint/observation in 1 dimension

This relationship then is used to find the support vector classifier

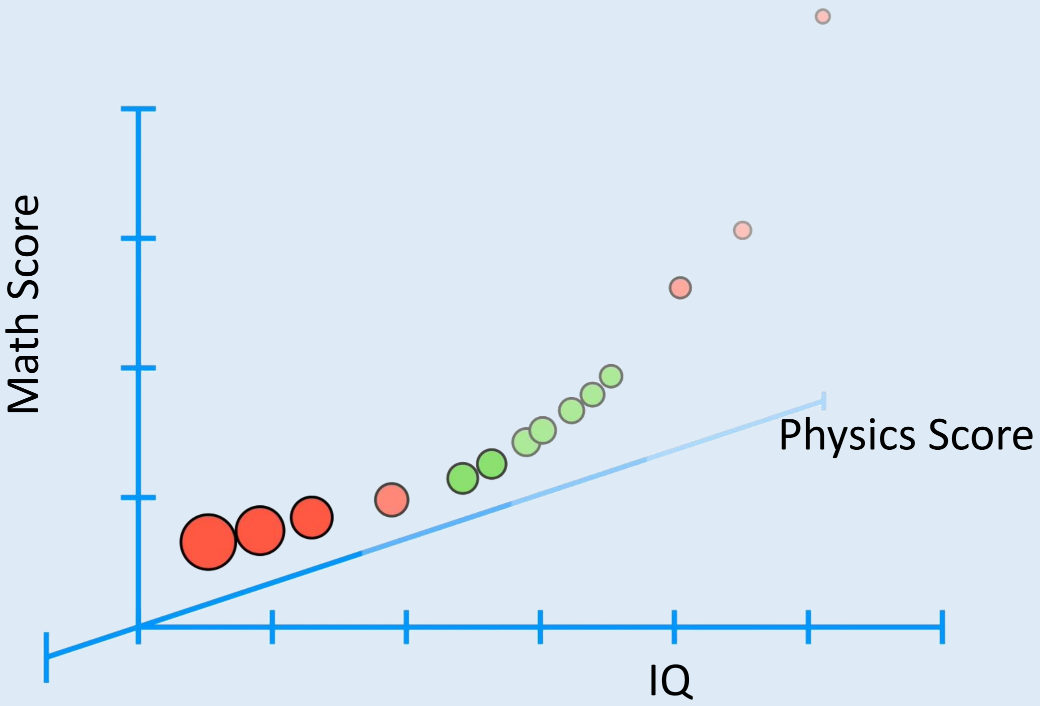


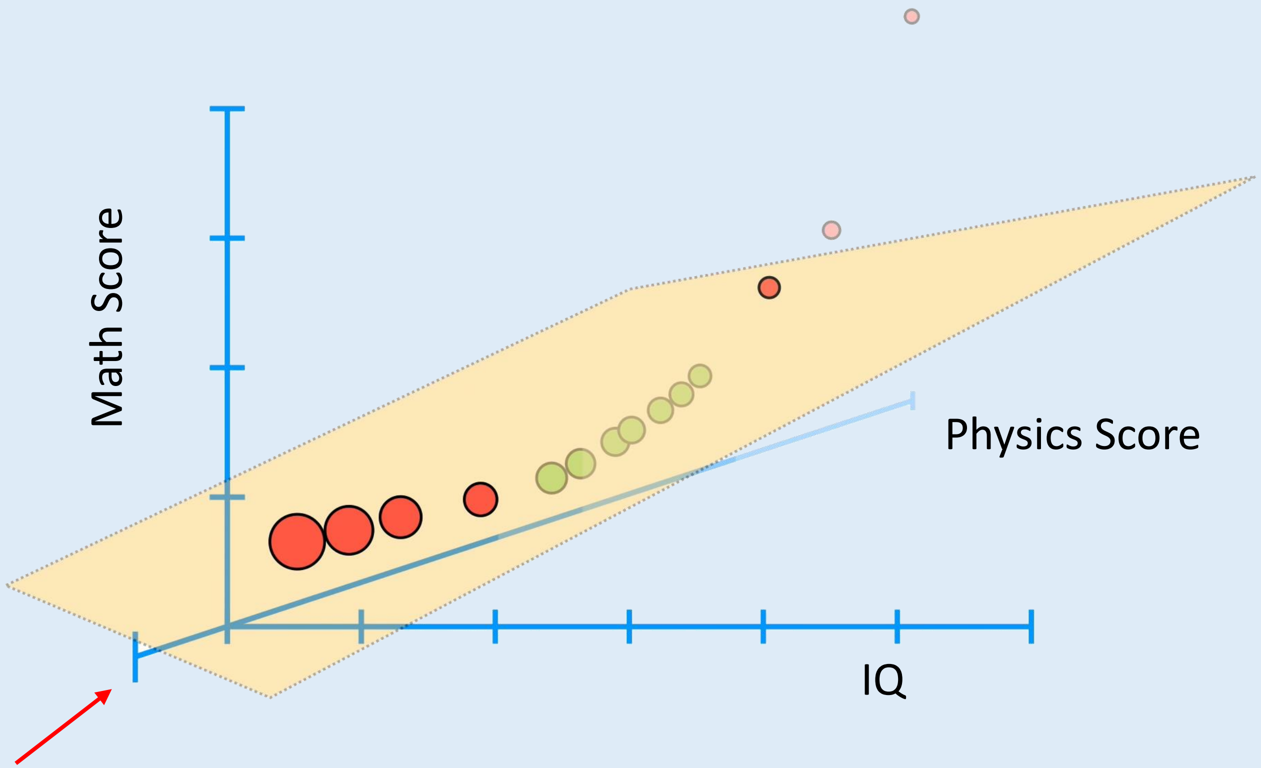
When d = 2, the Polynomial Kernel Function decides the relationship between each pair of datapoint/Observation in 2 Dimensional Data



This relationship then is used to find the support vector classifier

When d = 3, the Polynomial Kernel Function decides the relationship between each pair of datapoint/observation in 3 dimensional data





This relationship then is used to find the support vector classifier

For best results i.e. finding the right value of parameter d, cross validation can be used