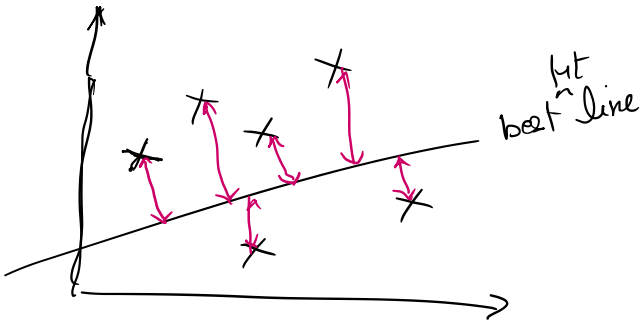
Support Vector for Regression :->

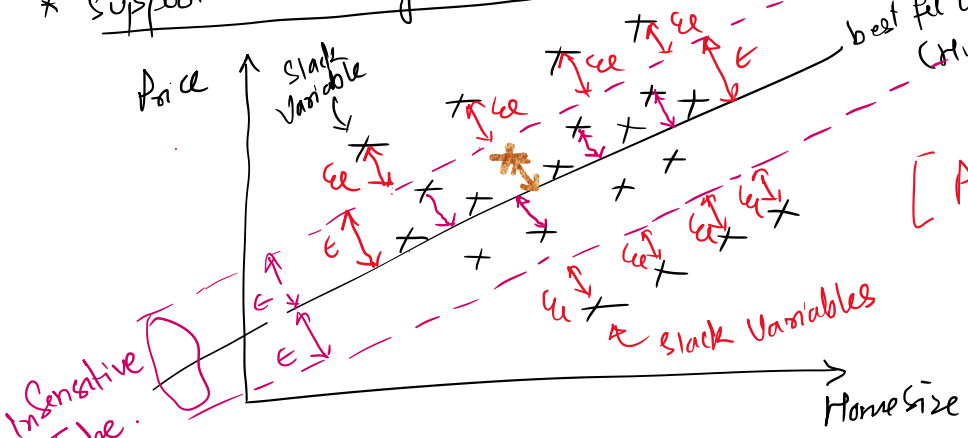
* In Linear Regression

Objective -> Minimize MSE
[Mean Squared Error]



Best Fit line -> gt is the line that gives Minimum MSE

* Support Vector Regression :->



[Data Points inside ϵ Insensitive Tube -> support Vector]

-> Here we are creating ϵ (Insensitive Tube) [Allowed Margin of Error]

* We can ignore the Error if difference between observed and predicted value is $\leq \epsilon$

* All the points inside ϵ Insensitive tube are not considered for Calculating Error.

* The data points outside the ϵ Insensitive tube are known as Slack Variables.

* For a point x_i within Insensitive (ϵ) tube

* For a point x_i within Insensitive (ϵ) tube

$|y_i - w^T x_i| \leq \epsilon$

y_i = observed value

x_i = predicted value.

Constraint ⇒ if $|y_i - w^T x_i| \leq \epsilon$ (y_i ↑, $w^T x_i$ ↑) we can say that prediction is good and we will not consider the difference in Error Calculation.

* For Slack Variable [points outside ϵ Insensitive tube]

we need to calculate distance of Slack Variable from Hyperplane

= distance bet^n Slack Variable + ϵ & marginal plane

= $\xi_i + \epsilon$.

Hinge Loss.

Objective ⇒

Cost F^n ⇒ Minimize $\frac{\|w\|}{2} + C \sum_{i=1}^{n} \xi_i$ (w, b)

Hyperparameter

Constraints $|y_i - w^T x_i| \leq \epsilon + \xi_i$ → for points outside ϵ Insensitive tube

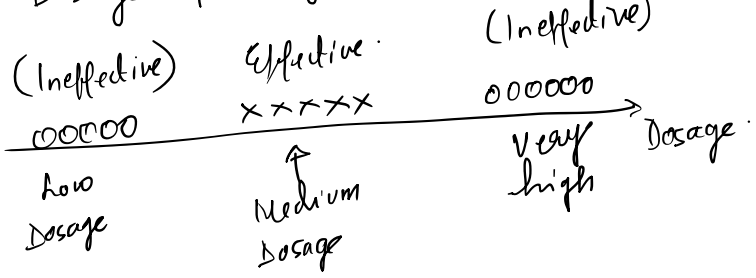
We have considered upto till Now Support Vector Classifier (Soft Margin)

[We were interested in finding Max Margin classifier]

* If there are high amount of Overlapping than soft Margin classifier will not be efficient

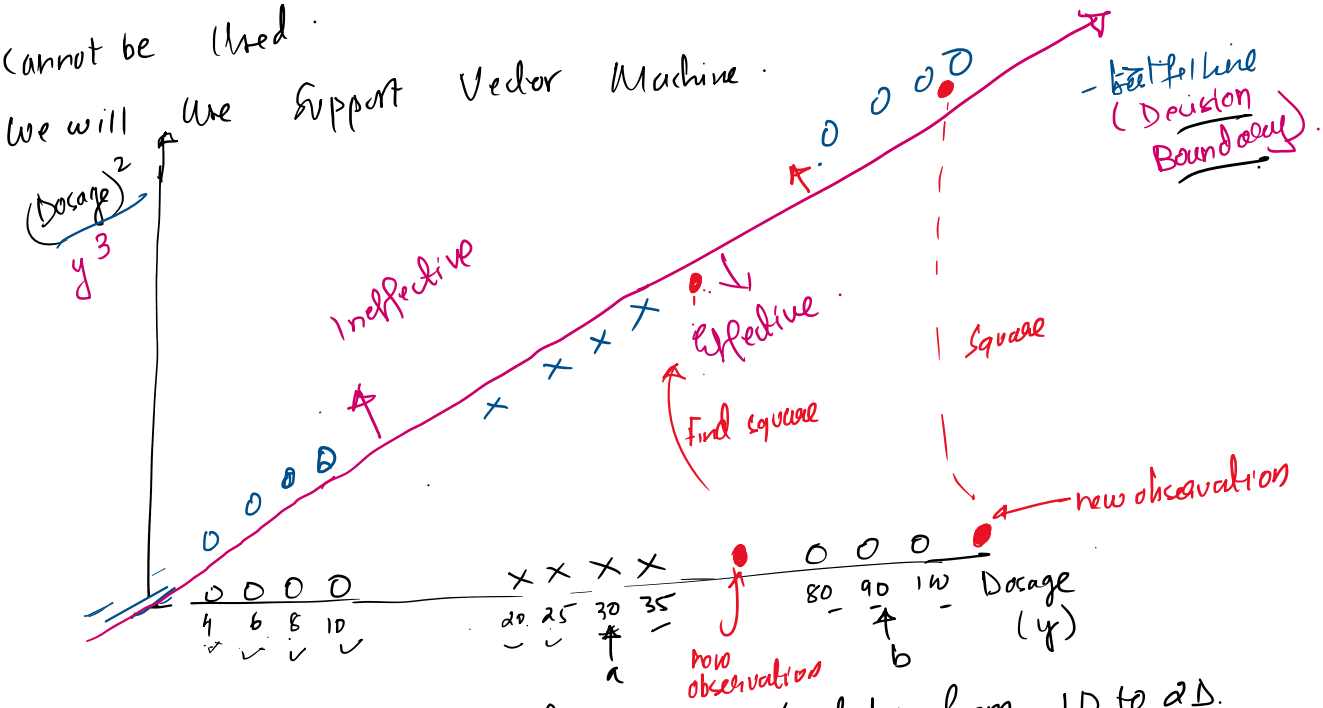
* In such case we will use Support Vector Machine.

Ex:- Consider the Dosage of Drug



→ Here we have lot of Overlapping so Maximum Margin classifier cannot be used.

↳ we will use Support Vector Machine.



* from above By increasing dimension of data from 1D to 2D. and by using Square (as function) we are able to separate the data

* from above "U" ——— "U" and by using Square (as function) we are able to separate the two points.

* We can also use it to predict for new observation
  → For new observation take square of the observation and chk if the square lies above (ineffective) or below (effective) the decision boundary.

SVM Main Idea →
* Start data with low dimension (Given).
* Now move data to Higher Dimension [Using Some function]
* Find Support Vector Classifier to classify the data into different classes.

Question Arises -> Why Square, why not Cube or Some Other function?

→ How to decide, how to transform.

* The Function Used be known as Kernel

* Types of Kernel function in SVM.
  (1) Polynomial Kernel
  (2) Radial Basis Kernel.

Note 1) If data points in 1D → Convert in 2D ⇒ line as Decision boundary
2D → Convert in 3D ⇒ Plane as Decision Boundary

2) If Data points in 1D → Point is Decision Boundary
2D → line is Decision Boundary
3D → Plane is Decision Boundary.

Polynomial Kernel

* It calculates Higher Dimensional Relationship between observations (data points).

* The Kernel that transforms 1D to 2D is Polynomial Kernel.

* It may look like $(a * b + r)^d$

where a and b are two different observations in dataset (any two data points).

r = coefficient of Polynomial.

d = degree of Polynomial.

* Here we use SVM with Polynomial Kernel to Compute Relationship bet^n observations in Higher dimension and then find Good classifier.

* Let a and b be two observations.
* Let $r = 1/2$ & $d = 2$ } Note ⇒ value of r and d is determined by Cross Validation.

$$\Rightarrow (a * b + r)^d = \left(a * b + \tfrac{1}{2}\right)^2$$

$$= \left(a * b + \tfrac{1}{2}\right) \cdot \left(a * b + \tfrac{1}{2}\right)$$

$$= a^2 b^2 + \tfrac{1}{2} ab + \tfrac{1}{2} ab + \tfrac{1}{4}$$

$$= ab + a^2 b^2 + \tfrac{1}{4} \Rightarrow \text{Polynomial.}$$

= can be written as dot product of

= can be written as dot product of

$= (a, a^2, \frac{1}{2}) \cdot (b, b^2, \frac{1}{2})$

For data point a — original value a — Higher dimension value of a — original value b — value of b in higher dimension

Dot product is sum of 1st term multiplied, and term multiplied and so on

Note: $\frac{1}{2}$ is third axis but value same so ignore.

So $(a + b \times \gamma)^d$ is used to get Higher dimension Rel^n between two data points a & b.

Let $a = 9$ $b = 14$

$\gamma = \frac{1}{2}$ $b = a$

$\Rightarrow$ Higher Dimension Relationship bet^n $a = 9$, $b = 14$

$\Rightarrow (9 + 14 + \frac{1}{2})^2 \Rightarrow 126.5^2 \Rightarrow 16002.25$

↑ Relationship Representation of $a = 9$, $b = 14$ in Higher Dimension

* We can find this Higher Dimension Relationship bet^n every pair of Data points.

VVImp → Kernel Trick

(1) Kernel function actually never does any transformation in Higher Dimension.

(2) Instead it calculates Relation bet^n observations and visualizes them in Higher Dimension.

visualizes them in Higher Dimension.

(3) Support Vector Classifier uses this visualization for classification.

(4) This is known as Kernel Trick.

$d=1$ ⇒ The Polynomial Kernel computes the relationship bet^n each pair of observation in 1D.

$d=2$ ⇒ " " " " " "

" " " " " " 2D.

Best value of d can be found by Cross Validation.

Note Since Polynomial Kernel is $(a*b+1/2)^2 = (a, a^2, 1/2)\cdot(b, b^2, 1/2)$

→ we actually do not have to Transform to understand higher dimension relationship.

→ All we need to do is calculate dot product bet^n each pair of point.

Radial Basis Kernel [Vimp]

* Works in Infinite Dimension.
* If there are lot of Overlapping in data points classification. then Support Vector Classifier cannot be used to linearly Separate.
* One Way to deal with overlapping data is to use sum with Radial Kernel.
* Radial Kernel Uses Radial Basis function!

$$e^{-\gamma \cdot (a-b)^2}$$

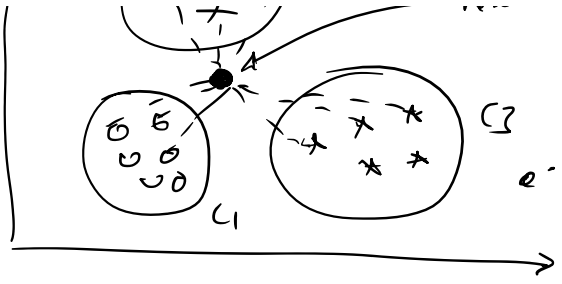
Here a and b are two observations.

$(a-b)^2 \Rightarrow$ diff b/w the measurement is squared. giving us the squared distance b/w two observation.

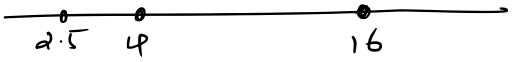
$\gamma \Rightarrow$ Scales the squared distance and thus Scales the influence.

* Since Radial Basis function finds Support Vector Classifier in Infinite Dimension it is not possible to visualise it.
* When applied to new observation, it behaves like weighted Nearest Model.
* Radial Kernel checks the influence of the Nearest

Influence of the Nearest Neighbour on New observation and accordingly makes classification.



* but see how radial Kernel determines how much influence each observation in training dataset has on classifying new observation.



but $a = 2.5$, $b = 4$ $\Rightarrow e^{-1(2.5-4)^2} \Rightarrow 0.11$ — (1)

$\gamma = 1$

$a = 2.5$, $b = 4$, $\gamma = 2$ $\Rightarrow e^{-2(2.5-4)^2} \Rightarrow 0.01$ } Thus γ scales the influence.

$a = 2.5$, $b = 16$, $\gamma = 1$ $\Rightarrow e^{-1(2.5-16)^2} \Rightarrow e^{-182.25} \approx$ close to zero

↳ (2)

From (1) & (2) if points are close enough then we have high influence.

and if points are far off then we have low influence.

* Thus to calculate influence b/t two data points, we can plug in the values in Radial Basis fn with appropriate γ.

Infinite Dimension.

* We get relationship in Infinite Dimension.

Not from Exam :⇒ Let a and b are two observations.

Let $\gamma = 1/2$

$$e^{-1/2\,(a-b)^2} = e^{-\frac{1}{2}(a^2+b^2-2ab)} = e^{-\frac{1}{2}(a^2+b^2)} \cdot e^{ab} \quad \rightarrow (I)$$

Taylor Series Expansion ⇒ [Allows any fn to split in infinite Sum]

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{\infty}(a)}{\infty!}(x-a)^{\infty}$$

Let $f(x) = e^x$.

$$e^x = e^a + \frac{e^a}{1!}(x-a) + \frac{e^a}{2!}(x-a)^2 + \cdots + \frac{e^a}{\infty!}(x-a)^{\infty}$$

Let $a = 0$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots \frac{x^{\infty}}{\infty!}.$$

$$\therefore \quad e^{ab} = 1 + \frac{ab}{1!} + \frac{a^2b^2}{2!} + \cdots \frac{a^{\infty}b^{\infty}}{\infty!}$$

$$e^{ab} = \left(1, \frac{1}{\sqrt{1!}}a, \frac{1}{\sqrt{2!}}a^2 \cdots, \frac{1}{\sqrt{\infty!}}a^{\infty}\right) \cdot \left(1, \frac{1}{\sqrt{1!}}b, \frac{1}{\sqrt{2!}}b^2 \cdots, \frac{1}{\sqrt{\infty!}}b^{\infty}\right) \quad \rightarrow (2)$$

from (1) & (2)

$$= e^{-\frac{1}{2}(a^2+b^2)} \left(1, \frac{a}{\sqrt{1!}}, \frac{a^2}{\sqrt{2!}} \cdots \frac{a^{\infty}}{\sqrt{\infty!}}\right) \cdot \left(1, \frac{b}{\sqrt{1!}}, \frac{b^2}{\sqrt{2!}} \cdots, \frac{b^{\infty}}{\sqrt{\infty!}}\right)$$

Let $\varphi = e^{\frac{-1}{2}(a^2+b^2)}$

$$= \left(\varphi, 1, \frac{a}{\sqrt{1!}}, \frac{a^2}{\sqrt{2!}} \cdots \frac{a^\infty}{\sqrt{\infty!}}\right) \cdot \left(\varphi, 1, \frac{b}{\sqrt{1!}}, \frac{b^2}{\sqrt{2!}}, \cdots \frac{b^\infty}{\sqrt{\infty!}}\right)$$

We can see Radial Kernel is equal to Dot product that has coordinate for infinite No of Dimension.