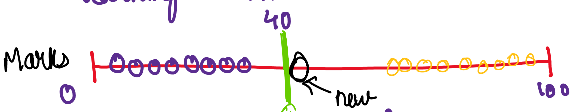
Support vector machine

Sunday, 31 March 2024 5:36 PM

SVMs act as a frontier that can segregate two classes in the best possible manner. The frontier that seperates two classes is also known as the decision boundary, and SVM helps us get an optimal decision boundary.

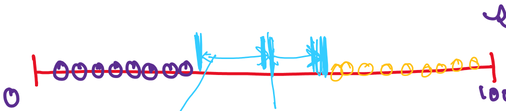
** The optimal decision boundary is known as hyperplane in machine learning context.



suppose, we took this as threshold.

O → section 8 AI
O → section 12 AI

This will be classified as section 8 but it is close to section 12. Therefore it is wrong

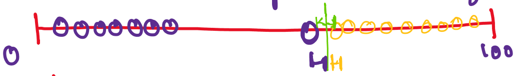


The threshold is half way between the observation

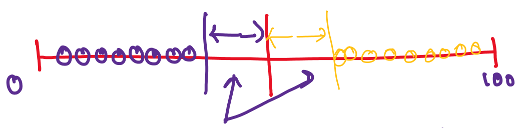
Margin: When we use the threshold that gives us the largest margin to make classifications.

Maximal Margin classifier

what if our training dataset look like this



So, maximum margin classifiers are super sensitive to outliers in the training data and that makes them pretty lame.



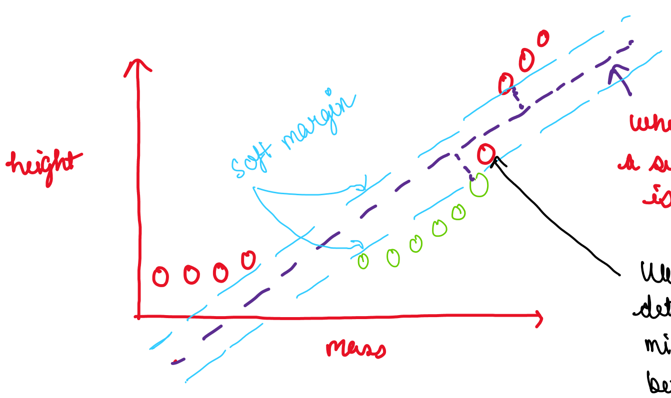
When we allow misclassifications, the distance between the observations and the threshold is called soft margin

We use cross validation to determine how many misclassification and observations to allow inside of the Soft margin to get the best classification.

We use soft margin to determine the location of a threshold

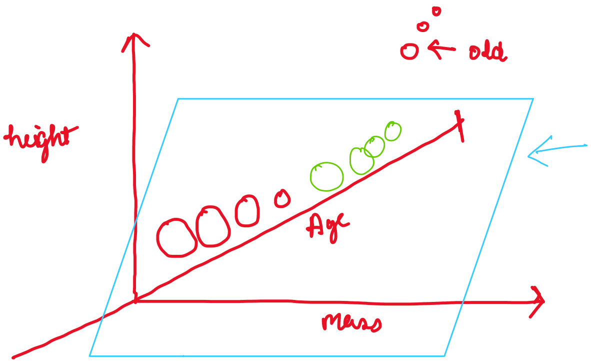
We are using Soft margin classifier aka a Support Vector classifier to classify the observations.

The name support vector classifier comes from the fact that the observations on the Edge and within the Soft margin are called Support Vectors.



when data is 2-Dimensional a support vector classifier is a line

We use Cross validation to determine that allowing this misclassification results in better classification in long run



when data is 3-dimensional the support vector Classifier forms a plane, instead of line

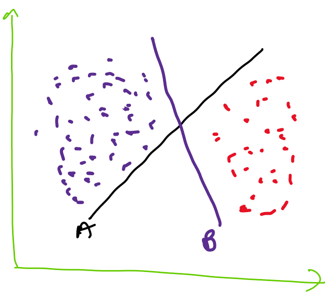
Data	Support vector Classifier	
1 Dimentional	Single point	(0-D)
2 Dimentional	line	(1-D)
3-D	plane	(2-D)
4 D or more	hyper plane	

Support Vector Machine acts as a frontier that can segregate two classes in the best possible manner

It works best when two classes need to be segregated, considering extreme characteristics of both the classes.

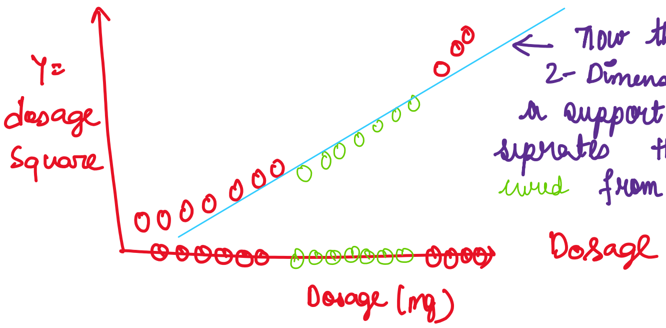
Why do we need an optimal decision boundary ?

An optimal decision boundary helps split the two classes, considering the Extreme points the Classes. It uses the data point that are closest to both the classes also known vectors, the optimal decision boundary is also known as a Hyperplane in machine learning



Workings of the SVM

SVMs, introduce linear discriminators/hyperplanes which maximizes the distance between the hyperplane and the "difficult points" close to hyper plane.



Now that the data are 2-Dimensional, we can draw a support vector classifier that seprates the people who were cured from the people who were not cured

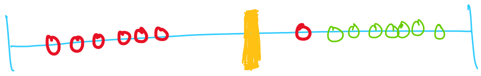
Step 1: Start with data in a relatively low dimension
Step 2: Move the data into a higher dimension.
Step 3: Find a Support Vector classifier that seperates the higher dimensional data into two groups.

How do we decide how to transform the data?

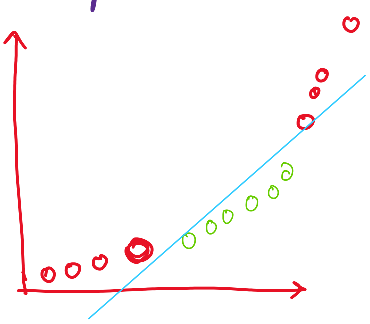
Support Vector machines uses something called Kernel Functions to systematically Find Support Vector Classifiers in higher dimension

Kernel Function systematically Finds Support Vector Classifiers in higher dimensions

When d = 1, the Polynomial Kernel computes the relationships between each pair of observations in 1-Dimension...



The polynomial kernel computes the 2-Dimensional relationships between each pair of observations.....



If d = 3, then we would get a 3rd dimension based on dosages^3 and the Polynomial Kernel computes the 3-Dimensional relationships between each pair of observations.

Another most commonly used kernel is the Radial Kernel, also known as the Radial Basis Function (RBF) Kernel

RBF Kernel Finds Support vector Classifiers in infinite dimensions

Radial kernel behaves like a weighted Nearest Neighbor model

Polynomial Kernel

$$(a \times b + r)^d$$

$$(a \times b + \tfrac{1}{2})^2 = (a * b + \tfrac{1}{2})(a * b + \tfrac{1}{2})$$

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

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

$$(a \times b + r)^d$$

a & b → refer to two different Observations in the dataset
r → determines the coefficient of the polynomial
d → degree of the polynomial

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

$r = 1$ $d = 2$

$$(a \times b + 1)^2 = (a \times b + 1)(a * b + 1)$$

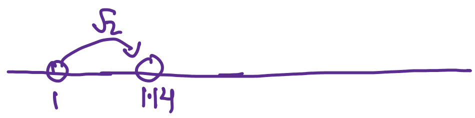
$$= 2ab + a^2b^2 + 1$$

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

ignore because it is a constant value

move by the factor of $\sqrt{2}$

Square



r & d are determine using cross validation.

Radial Kernel $e^{-r(a-b)^2}$

One way to deal with overlapping data is to use a support vector Machine with Radial Kernel

$$e^{-r(a-b)^2}$$

For a new observation The Radial Kernel behaves like a weighted Nearest Neighbor model

In other words, the closest observation (aka the nearest neighbors) have a lot of influence on how we classify the new observation

$e^{-\gamma(a-b)^2}$ $a \rightarrow$, $b \rightarrow$ { refer to two different Dosage measurement }

$\gamma \rightarrow$ (gamma), which is determined by Cross validation, scales the Squared distance and thus, it scales the influence.

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

$$e^{-(2.5-4)^2} = e^{-(-1.5)^2} = e^{-2.25} = 0.11$$

when $\gamma = 1$

far observation

$$e^{-(2.5-16)^2} = e^{-(-13.5)^2} = e^{-182.25} = \text{A Number Very close to zero}$$

$$a^1b^1 + a^2b^2 + a^3b^3 + \cdots + a^\infty b^\infty = (a, a^2, a^3, \ldots, a^\infty).(b, b^2, b^3, \ldots$$

Radial Kernel

let $\gamma = 1/2$

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

$$= e^{-1/2(a^2+b^2)} \underline{e^{ab}}$$

$\downarrow$

Talor series Expansion.

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

$$e^x = e^a + \frac{e^a}{1!}(x-a) + \frac{e^a}{2!}(x-a)^2 + \frac{e^a}{3!}(x-a)^3 + \cdots + \frac{e^a}{\infty!}(x-a)^\infty$$

The definition of the taylor series says that a can be any value as long as $f(a)$ exists

since $e^0 = 1$, e^0 exists

we set $a = 0$

$$e^x = e^0 + \frac{e^0}{1!}(x-0) + \frac{e^0}{2!}(x-0)^2 + \frac{e^0}{3!}(x-0)^3 + \cdots + \frac{e^0}{\infty!}(x-0)^\infty$$

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1x^3}{3!} \cdots + \frac{1}{\infty!}x^\infty$$

$$e^{ab} = 1 + \frac{1}{1!}ab + \frac{1}{2!}(ab)^2 + \frac{1}{3!}(ab)^3 + \cdots + \frac{1}{\infty!}(ab)^\infty$$

$$= a^0 b^0 + \frac{1}{1!}ab + - - + \frac{1}{\infty!}(ab)^\infty$$

$$\left(a^0, a^1, \frac{1}{\sqrt{2!}}a^2, \frac{1}{\sqrt{3!}}a^3 --- \frac{1}{\sqrt{\infty!}}a^\infty\right)\left(b^0, b^1, \frac{1}{\sqrt{2!}}b^2 + \frac{1}{\sqrt{3!}}b^3 + --- \frac{1}{\sqrt{\infty!}}b^\infty\right)$$

$$RBF = e^{-1/2(a^2+b^2)}e^{ab}$$

$$= e^{-\frac{1}{2}(a^2+b^2)}\left(a^0, a^1, \frac{1}{\sqrt{2!}}a^2, \frac{1}{\sqrt{3!}}a^3 --- \frac{1}{\sqrt{\infty!}}a^\infty\right)\left(b^0, b^1, \frac{1}{\sqrt{2!}}b^2 + \frac{1}{\sqrt{3!}}b^3 + --- \frac{1}{\sqrt{\infty!}}\right)$$

assume S

$$= S\left(a^0, a^1, \frac{1}{\sqrt{2!}}a^2, \frac{1}{\sqrt{3!}}a^3 --- \frac{1}{\sqrt{\infty!}}a^\infty\right)\left(b^0, b^1, \frac{1}{\sqrt{2!}}b^2 + \frac{1}{\sqrt{3!}}b^3 + --- \frac{1}{\sqrt{\infty!}}b^\infty\right)$$

$$= \left(\sqrt{s}a^0, \sqrt{s}a^1, \frac{\sqrt{s}1}{\sqrt{2!}}a^2, \frac{\sqrt{s}1}{\sqrt{3!}}a^3 --- \frac{\sqrt{s}1}{\sqrt{\infty!}}a^\infty\right)\left(\sqrt{s}b^0, \sqrt{s}b^1, \frac{\sqrt{s}1}{\sqrt{2!}}b^2 + \frac{1\sqrt{s}}{\sqrt{3!}}b^3 + --- \frac{1\sqrt{s}}{\sqrt{\infty!}}b^\infty\right)$$

Therefore it is infinite dimensions

Linear Kernels

[3, 6] and [2, 5]

$\Rightarrow 3*2 + 6*5 = 36$

$f(x) = B_0 + \text{sum}(a_i * (x, x_i))$

Polynomial Kernels

$(a \times b + r)^d$

Radial Kernels

$e^{-r(a-b)^2}$