Support Vector Machine

Support vectors are data points that lie closest to decision surface

Points that are most difficult to classify

SVMs maximize the margin thus

Error = Classification Error + Margin Error

SVM helps to minimize classification error and margin error.

As we increase margin few points may become mis-classified so we include those in the error.

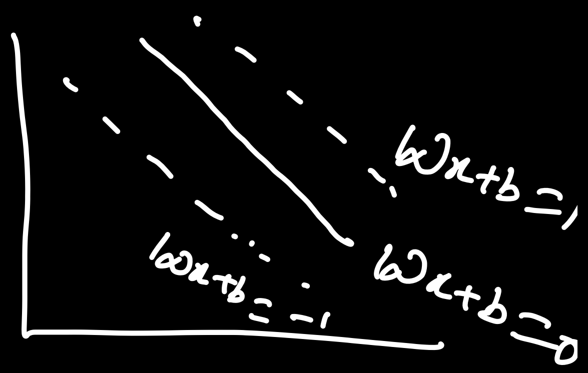
We want to turn margin to error so that it can be minimized by Gradient Descent.

We want a function that gives
small error when large margin
large error when small margin

Goal → Model with large margin

$$\text{Margin} = \frac{2}{|w|}$$

$$= \frac{2}{\sqrt{w_1^2 + w_2^2}}$$



$wx+b=0$

$w_1x_1 + w_2x_2 + b = 0$

Distance between ||ly lines is equal to distance of a

Eqn of || line slope same

pt to a line
perpendicular distance

$w_1x_1 + w_2x_2 + b = 1$

$w_2x_1 + w_2x_2 + b = -1$

$\left(-\frac{b}{w_1}, 0\right)$ ⊥ distn $w_1x_1 + w_2x_2 + b = 1$

$$d_1 = \frac{\left| w_1 \cdot \frac{-b}{w_1} + w_2 \cdot 0 + b - 1 \right|}{\sqrt{w_1^2 + w^2}} = \frac{1}{|w|}$$

llly $d_2 = \frac{1}{|w|}$ $\Rightarrow d = \frac{2}{||w|}$

Thus

Margin is $\frac{2}{|w|} = \frac{2}{\sqrt{w_1^2 + w_2^2}}$

Now

Error $= \sqrt{w_1^2 + w^2}$

Now Margin $\propto \frac{1}{\text{Error}}$

Thus

Large Error ⇒ Small Margin

Small Error $\Rightarrow$ Large Margin

Ex 1

$w = (3,4) \quad b = 1$

$\Rightarrow \quad 3x_1 + 4x_2 + 1 = 0$

other lines $\quad 3x_1 + 4x_2 + 1 = 1$

$3x_1 + 4x_2 + 1 = -1$

Error $= \|w\|^2 = 3^2 + 4^2 = 25$

Margin $= 2/\|w\| = 2/5$

Ex 2

$w = (6,8) \quad b = 2 \qquad \Big| \; *2$

$\Rightarrow \quad 6x_1 + 8x_2 + 2 = 0$

other lines

$6x_1 + 8x_2 + 2 = 1$

$6x_1 + 8x_2 + 2 = -1$

Error $= \|w\|^2 = 6^2 + 8^2 = 100$

Margin $= 2/10 = 1/5$

Now we have two models with same boundary line but different margins

$M1 = 2/5$ $\quad$ $M2 = 1/5$

$E1 = 25$ $\quad$ $E2 = 100$

$\uparrow M \downarrow E$ $\quad$ $\downarrow M \uparrow E$

Margin Error

is norm of vector W → squared

Exactly same as given by Regularization term in L2 Regularization

ERROR FUNCTION

Error = Classification Error + Margin Error

Minimize using Gradient Descent

The C Parameter

Error = C * classification error + Margin Error

If C is large then we are concerned with classification error
we want all points to be classified correctly (Medical model)

If C is small then it is mainly margin error
(maybe few classification error)

We can use Grid Search for value of C

POLYNOMIAL KERNEL

Kernel trick - line is not enough

Kernel means set of functions that will help to separate

2D → 5D

$(x, y) \rightarrow (x, y, x^2, xy, y^2)$

project 2D $\Leftarrow$ 4D hyperplane can separate

We add dimensions to the data and find higher dimension surface that separates and project down to get curves

Poly

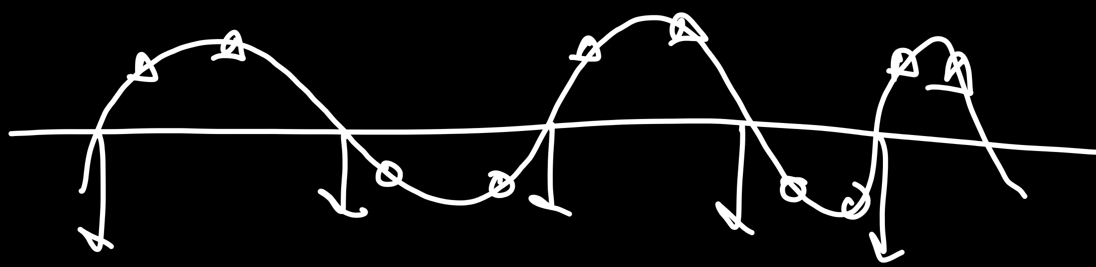
$$x, y$$
$$x^2, y^2, xy$$
$$x^3, x^2y, xy^2, y^3$$

RBF

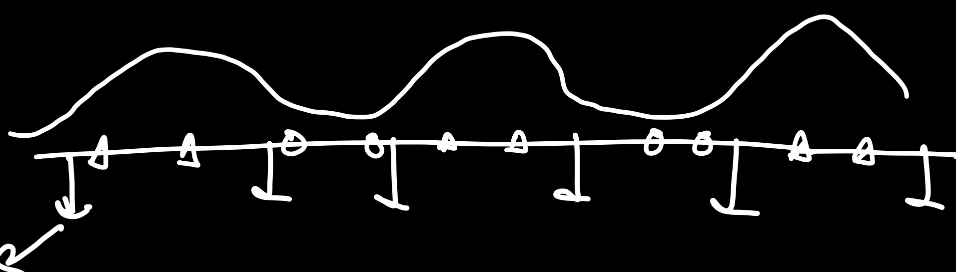
Radial Basis Function Kernel

There is no line that separates but we want a curve that has different heights for different categories.

When we move the points on this curve, the points can be separated with a line.



Where the line cuts the curve
is Model
and project is down

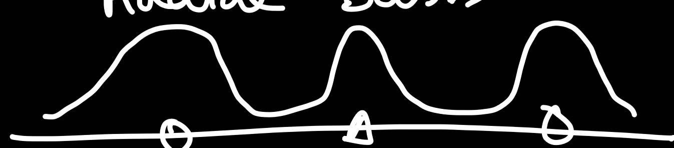


Model

Q How to find mountain range that is high for category and low for other

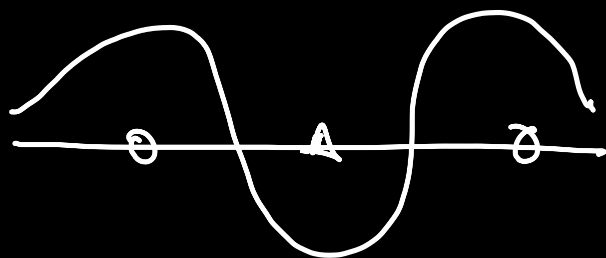
Ans
Draw one mountain range for each point that is high and then combine these

One mountain range is called
Radial Basis Function



Radial Basis Fn

Multiply +1 , -1 , +1 add

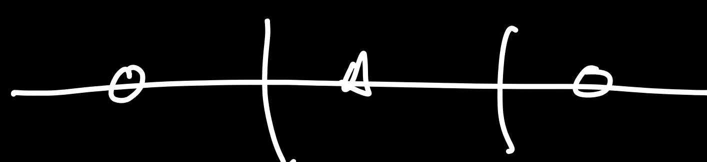


Now shift the points

We can find a line that separates these

We can find model when line and curve meets

and project down to get model

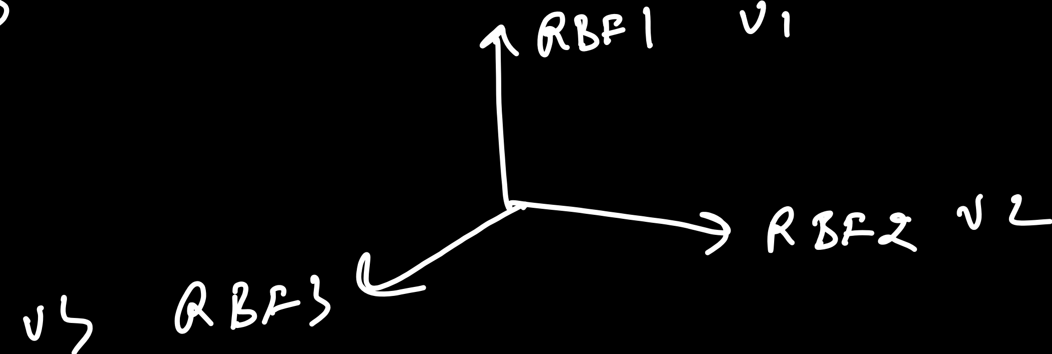


We can multiply Radial Basis Function with any number => weights

Construct a vector with length equal to no. of radial basis functions, each representing height of each function.

$$\underset{\text{RBF 1}}{\begin{bmatrix} 1 \\ 0.08 \\ 0.02 \end{bmatrix}} = v_1 \qquad \underset{\text{RBF 2}}{\begin{bmatrix} 0.08 \\ 1 \\ 0.06 \end{bmatrix}} = v_2 \qquad \underset{\text{RBF 3}}{\begin{bmatrix} 0.02 \\ 0.06 \\ 1 \end{bmatrix}} = v_3$$

Now how to find linear combination of these vectors that separates points



We have 3-basis vector since each one will have value 1 and other nearly zero

We can project each point to plane

Get a plane that separates

Co-eff of plane that separates give

linear combination of Radial Basis Function

In Higher Dimensions

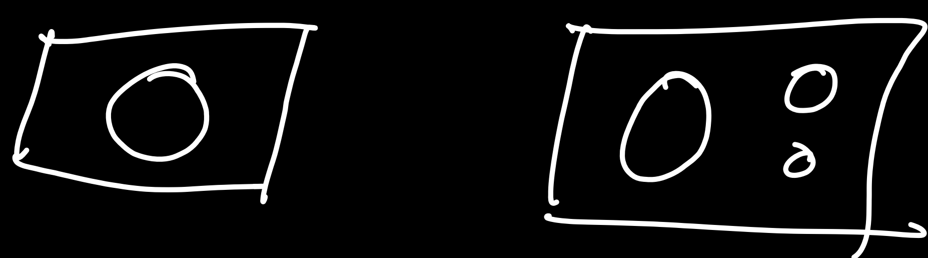
Mountain or Radial Basis Function is Gaussian Paraboloid

We can shift points over paraboloid

We can find a plane that cuts paraboloid

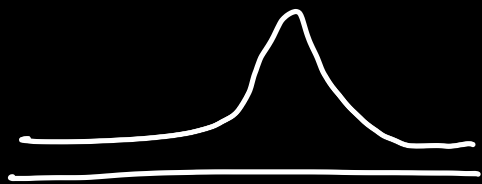
Since plane cuts paraboloid at circle circle will become boundary

Multiple Paraboloid may be



How to decide Radial Basis Function

γ

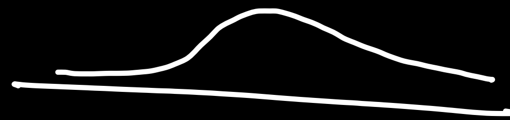


Large γ

Pointy Mountain

Narrow Curve

<u>Overfit</u>

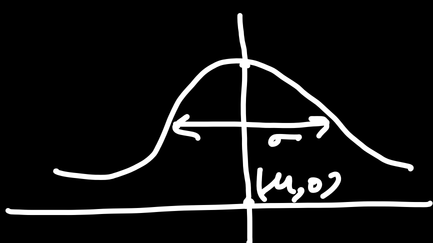


Small γ

Wide Curve

<u>Underfit</u>

γ = Gaussian / Normal Distribution



$$y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$ at origin

In general

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ is mean $\Rightarrow$ center of curve

σ is standard deviation $\rightarrow$ width of curve

$$\gamma \propto \frac{1}{\sigma}$$

$$\gamma = \frac{1}{2\sigma^2}$$

Small σ $\Rightarrow$ Large γ $\Rightarrow$ Narrow Curve

Large σ $\Rightarrow$ Small γ $\Rightarrow$ Wide curve