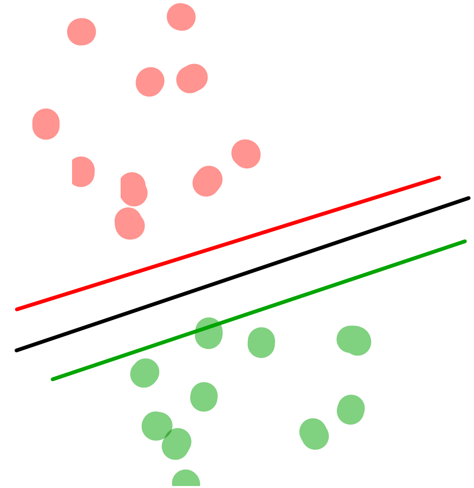
SVM-2

Recap

① Hard Margin Classifier

100% linearly separable.

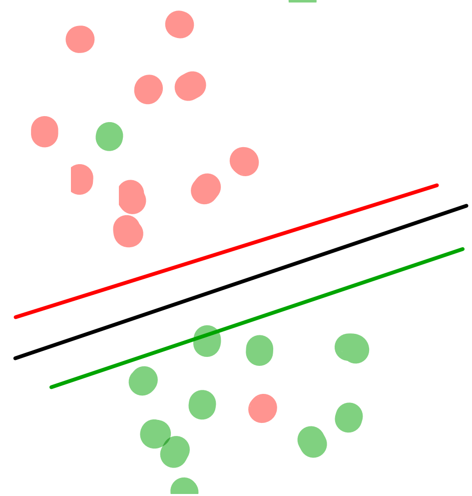


② Soft margin classifier.

Almost linearly separable

└ some errors.

Loss → HINGE LOSS.



3

primal

$$\arg\min_{w} \frac{\|w\|}{2} + \frac{C}{n}\sum_{i=1}^{n} \varepsilon_i \quad s.t \quad (w^T x + b)y \geq 1 - \varepsilon_i$$

4

dual

$$\arg\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \, y_i y_j \, x_i^T x_j$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \qquad 0 \leq \alpha \leq C$$

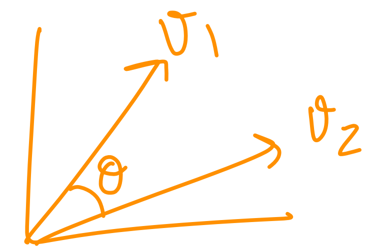
$$\Rightarrow \arg\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \underbrace{\alpha_i\alpha_j}_{\text{parameter}} \; \underbrace{y_i y_j}_{\text{label actual.}} \; \boxed{x_i^T x_j}$$

Mathematical function.

dot product
↳ Scalar

→ representing similarity b/w x_i & x_j

Cosine similarity



Kernel function: $K(x_i, x_j)$

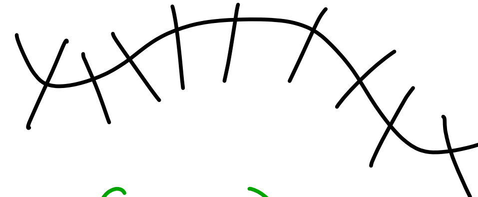
→ should represent similarity b/w any 2 vector.

Biotechnology

P1 C

P2 NC

P3 NC



100

↳ DNA

↳ 6 Can

$sim(P_1, P_2)$

$sim(P_1, P_3)$

$sim(P_2, P_3)$

Cancer

Noncancer

Amazon

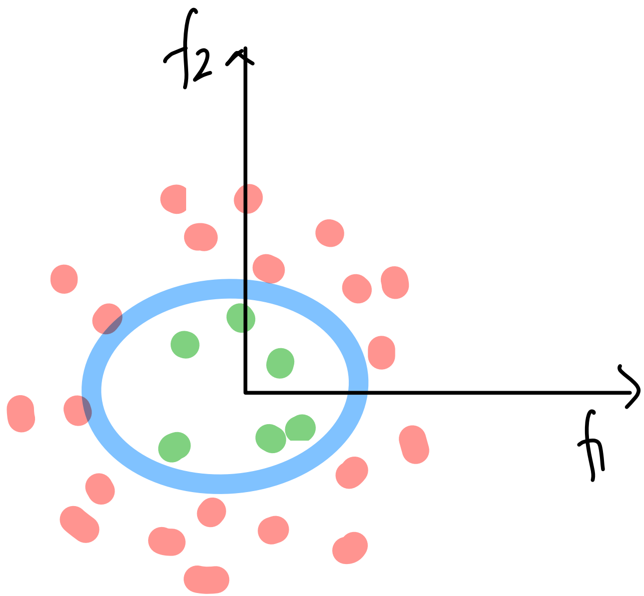
C1

C2

C3

This concept was applicable is almost all fields of study.

Kernel Function : Polynomial Kernel · $(C + x_i^T x_j)^n$

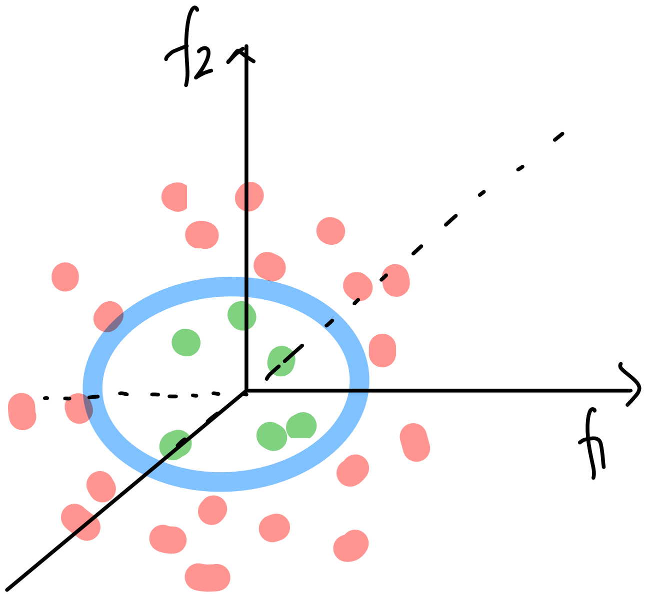


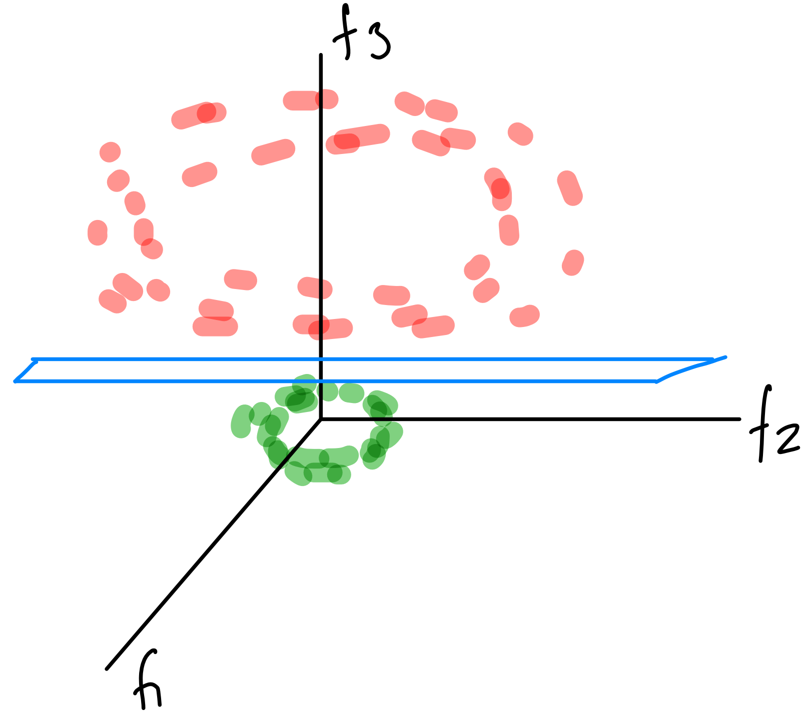
f_1, f_2

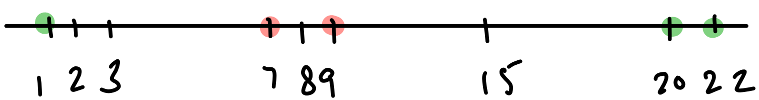
polynomial features

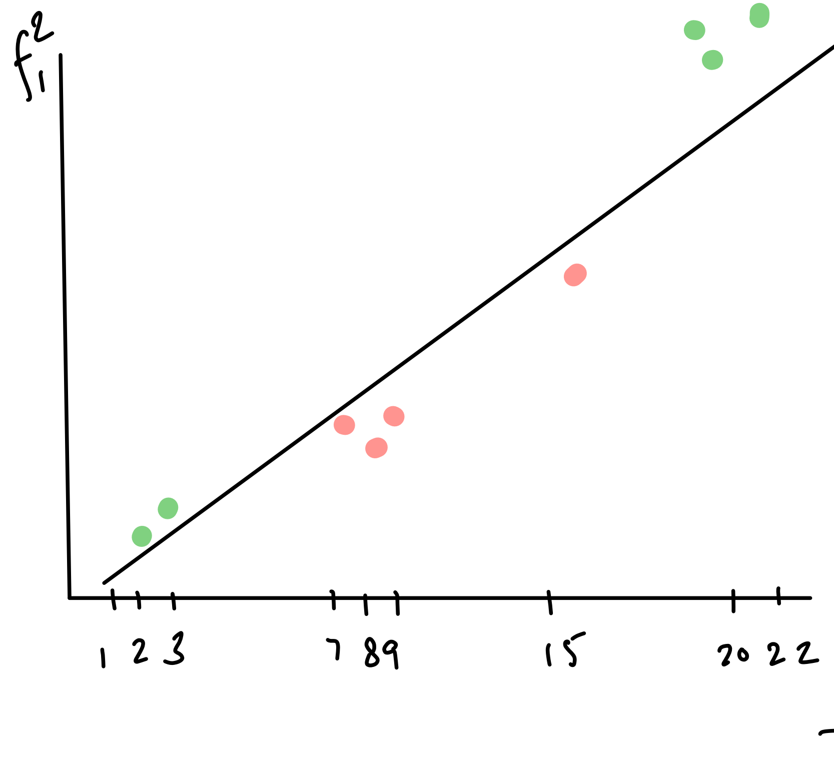
f_1^2, f_2^2

$f_1^2 + f_2^2 + \dots = 0$


f2
f1


f3
f2
f1


1 2 3
7 8 9
15
20 22


f_1^2
1 2 3
7 8 9
15
20 22
f_1

Polynomial Kernel $\Rightarrow$ $\left(\underbrace{C}_{\text{Const.}} + x_1^T x_2\right)^{n}$ degree

Const = 1, Degree = 2

$$K(x_1, x_2) = \left(1 + x_1^T x_2\right)^2$$ Quadratic Kernel

Const = 1 Degree = 3

$$K(x_1, x_2) = \left(1 + x_1^T x_2\right)^3$$ Cubic Kernel.

Kernelisation

d dim $\xrightarrow[\text{implicitly}]{}$ d' dims

$d' >>> d$

$2d$

$x_1 = \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix}$

$x_2 = \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}$

$d = 2$

$$K(x_1, x_2) = (1 + x_1^T x_2)^2$$

$$= \left(1 + \begin{bmatrix} x_{11} & x_{12} \end{bmatrix} \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}\right)^2$$

$$= (1 + x_{11}x_{21} + x_{12}x_{22})^2$$

2 dim $\xrightarrow[\text{implicitly}]{}$ 6 dim

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

$$= (1 + x_{11}x_{21} + x_{12}x_{22})^2$$

$$= \left(1 + x_{11}^2 x_{21}^2 + x_{12}^2 x_{22}^2 + 2x_{11}x_{21} + 2x_{12}x_{22} + 2x_{11}x_{21}x_{12}x_{22}\right)$$

$$x_1'^T = [1, x_{11}^2, x_{12}^2, \sqrt{2}x_{11}, \sqrt{2}x_{12}, \sqrt{2}x_{11}x_{12}]$$

$$x_2'^T = [1, x_{21}^2, x_{22}^2, \sqrt{2}x_{21}, \sqrt{2}x_{22}, \sqrt{2}x_{21}x_{22}]$$

$\underline{\underline{d = 6}}$

dot product $(x_1', x_2') \Rightarrow K(x_1, x_2)$

same result

$(x_1, x_2) \Rightarrow (x_1', x_2')$

2d 6d

Kernel Trick

$2d \xrightarrow[\text{Quadratic}]{} 6d$

$2d \xrightarrow[\text{Cubic}]{} 10d$

$3d \xrightarrow[\text{Quad}]{}$

$3d \xrightarrow[\text{Cubic}]{}$

$$[1 + x + y + z]^2$$

SVM → Kernel Trick

Log Reg + Poly features + Reg

V/S

SVM + $\underbrace{\text{Poly Kernel}}_{\text{Kernel}}$ + Reg

Radial Basis Function (RBF)

$$d = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + \ldots}$$

d dims $\rightarrow$ ∞ dims

$$\boxed{K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}}$$

$\|x_i - x_j\|^2 \rightarrow$ Euclidean distance.

$\sigma \rightarrow$ hyperparameter.

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

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

} → Normal Distribution
Gaussian

$$\arg\max_{\alpha} \sum \alpha - \frac{1}{2}\sum\sum \alpha_i \alpha_j y_i y_j \boxed{RBF}$$

$d(x_1, x_2) = 0.2$

$d(x_1, x_3) = 0.6$

$s = 0.6$

−10 10

$\sigma = 0.6$

$RBF(x_1, x_2) = 0.9$

$RBF(x_1, x_3) = 0.6$

σ_1

10

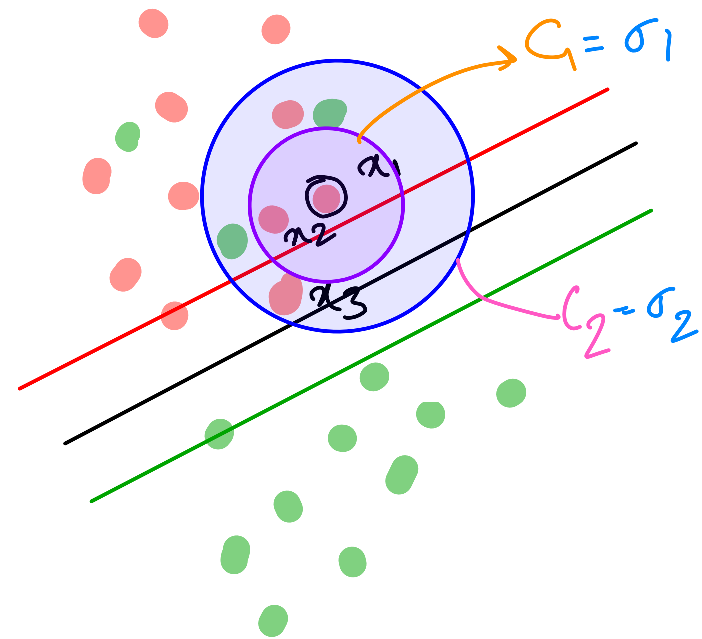
$\sigma = 1$

$RBF(x_1, x_2) = 0.95$

$RBF(x_1, x_3) = 0.9$

σ_2

KNN



$\sigma \uparrow \Rightarrow$ Underfitting

$\sigma \downarrow \Rightarrow$ Overfitting

KNN

$k \uparrow \rightarrow$ Underfit

$k \downarrow \rightarrow$ Overfitting.

d dim $\rightarrow$ ∞ dims

$$e^{-\frac{\|x_1 - x_j\|^2}{2\sigma^2}}$$ } $\rightarrow$ Euclidean distance

$\rightarrow$ 10000 dims $\rightarrow$ RBF $\rightarrow$ Euclidean Distance fails in higher dimension.

Curse of Dimensionality

Impact of Outlier on SVM

farther away from their class.

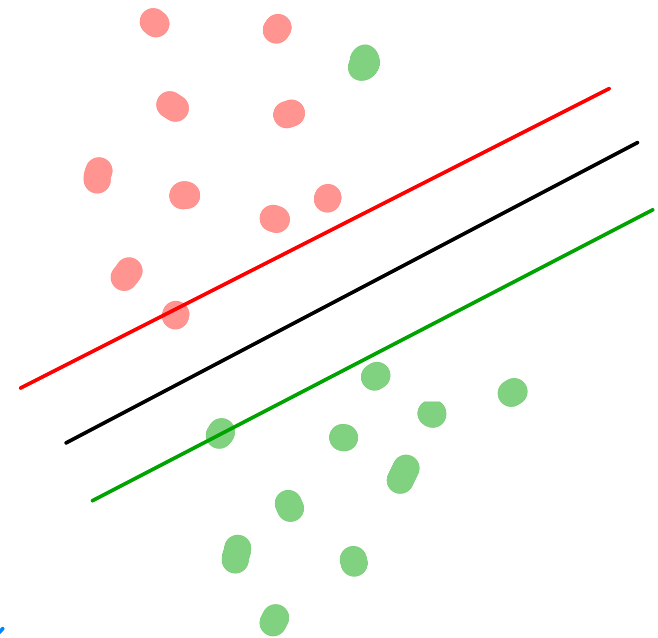
SVM + Kernel

1. Svm + Kernel
   f(distance)
   RBF

   Impact of outlier ↑↑↑

2. SVM + Kernel
   ~~f(distance)~~
   Poly

   Impact of outlier ↓↓↓



Why SVMs are not used?

* Computationally very expensive
* Training time ↑
* Time Complexity $O(n^2)$ ↑↑↑

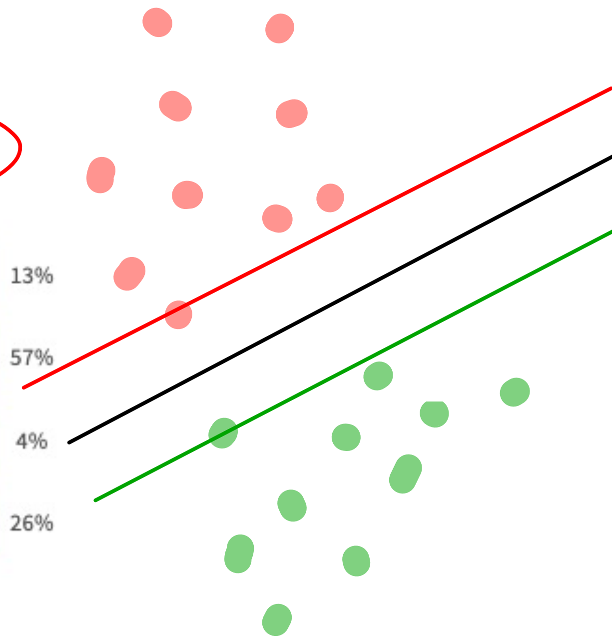
Quiz time!

Quiz Ended!

Which of the following are support vectors?

23 users have participated

A	Points which are within the margin	13%
B	Points which lie on +ve/-ve hyperplane (π^+/π^-)	57%
C	Points which are misclassified	4%
D	All of the above	26%



Quiz time!

Quiz Ended!

We have 100 datapoints out of which 5 are Support Vectors, then which is True:

5 $\alpha > 0$

95 $\alpha = 0$

21 users have participated

A	α >0 for 95 datapoints	19%
B	α <0 for 95 datapoints	14%
C	α = 0 for 5 datapoints	24%
D	α >0 for 5 datapoints	43%

Quiz time!

Quiz Ended!

What of the following statement(s) is/are true about Kernel in SVM? Statement 1: Kernel function map low dimensional data to high dimensional space
Statement 2: It's a similarity function

22 users have participated

A	Statement 1	5%
B	Statement 2	0%
C	Statement 1 and 2	95%
D	None of the above	0%

tgt ↓ test ↑

Quiz time!

Quiz Ended!

Both RBF and kNN are similar. Will they have similar run time as well?

19 users have participated

A	Yes	32%
B	No	63%
C	Don't know	5%

Not

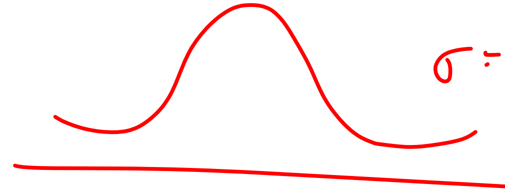
Quiz time!

Quiz Ended!

if σ decreases, what happens to RBF curve?

19 users have participated

A	Curve gets thinner	68%
B	Curve gets thicker	26%
C	No change	5%



$\sigma = 1$

$\sigma = 0.5$