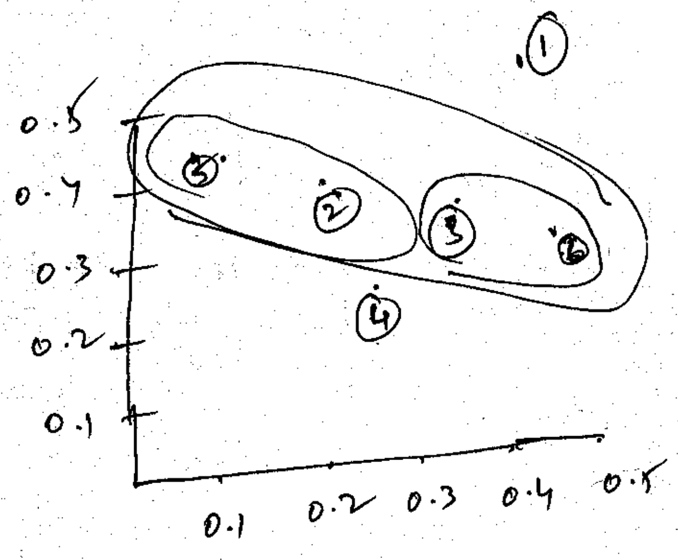
Clustering

- leaders
- Kmeans
- Kmeans ++

Single linkage

	x	y
P_1	0.40	0.53
P_2	0.22	0.38
P_3	0.35	0.32
P_4	0.26	0.19
P_5	0.08	0.41
P_6	0.45	0.30



$$d(P_1,P_2) = \sqrt{(x-a)^2 + (x-b)^2}$$

$$d[(0.40,0.53),(0.22,0.38)] = 0.23$$

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0					
P_2	0.23	0				
P_3	0.22	0.15	0			
P_4	0.37	0.20	0.15	0		
P_5	0.34	0.14	0.28	0.29	0	
P_6	0.23	0.25	0.11	0.22	0.39	0

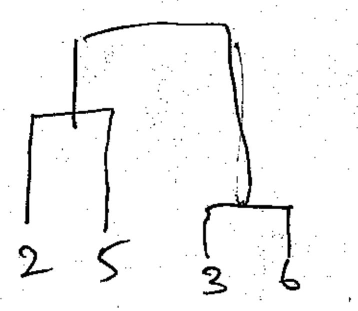
$(P_3,P_6,\ P_1) = \min(0.22, 0.23)$

$(P_3,P_6,\ P_2) = \min(0.15, 0.25)$

$(P_3,P_6,\ P_4) = \min(0.15, 0.22)$

$(P_3,P_6,\ P_5) = \min(0.28, 0.39)$

	P_1	P_2	$(P_3 P_6)$	P_4	P_5
P_1	0				
P_2	0.23	0			
$P_3 P_6$	0.22	0.15	0		
P_4	0.37	0.20	0.15	0	
P_5	0.34	0.14	0.28	0.29	0



	P_1	$P_2 P_5$	$P_3 P_6$	P_4
P_1	0			
$P_2 P_5$	0.23	0		
$P_3 P_6$	0.22	0.15	0	
P_4	0.37	0.20	0.15	0

$(P_1, P_2 P_5) = (0.23, 0.34)$
$= 0.23$

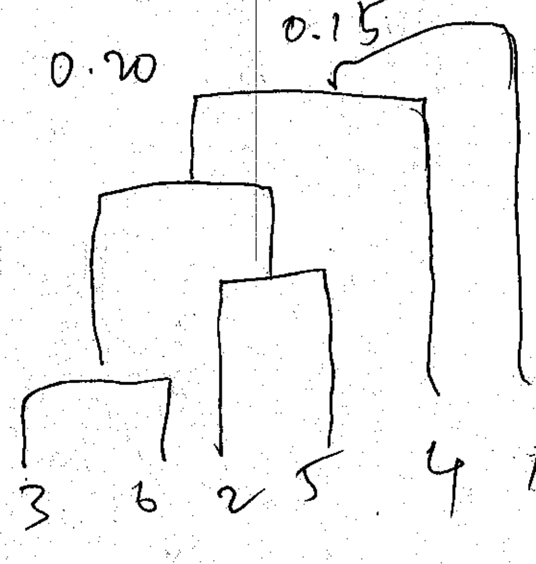
$(P_2(P_3, P_6)),$
$P_5(P_3 P_6)) = (0.15, 0.28)$
$= 0.15$

$(P_2 P_4\ P_5 P_4)\ (0.20, 0.29)$
$= 0.20$

	P_1	$(P_2 P_5)(P_3 P_6)$	P_4
P_1	0		
$(P_2 P_5, P_3 P_6)$	0.22	0	
P_4	0.37	0.15	0

$\text{dist}((P_2, P_5), P_1), ((P_3, P_6), P_1)$
0.22
$= 0.23$
$= 0.22$

$(P_2 P_5)\ P_4 \quad (P_3 P_6)\ P_4$
0.20 0.15



	P_1	$(P_2 P_5)(P_3 P_6)$	P_4
P_1			
$P_2 P_5$ $P_3 P_6$			
P_4			

Kmeans ++

K mediods

Kmeans ++

1. Select first random point
2. Select farthest random point
3. Select the farthest random point among the other two points.

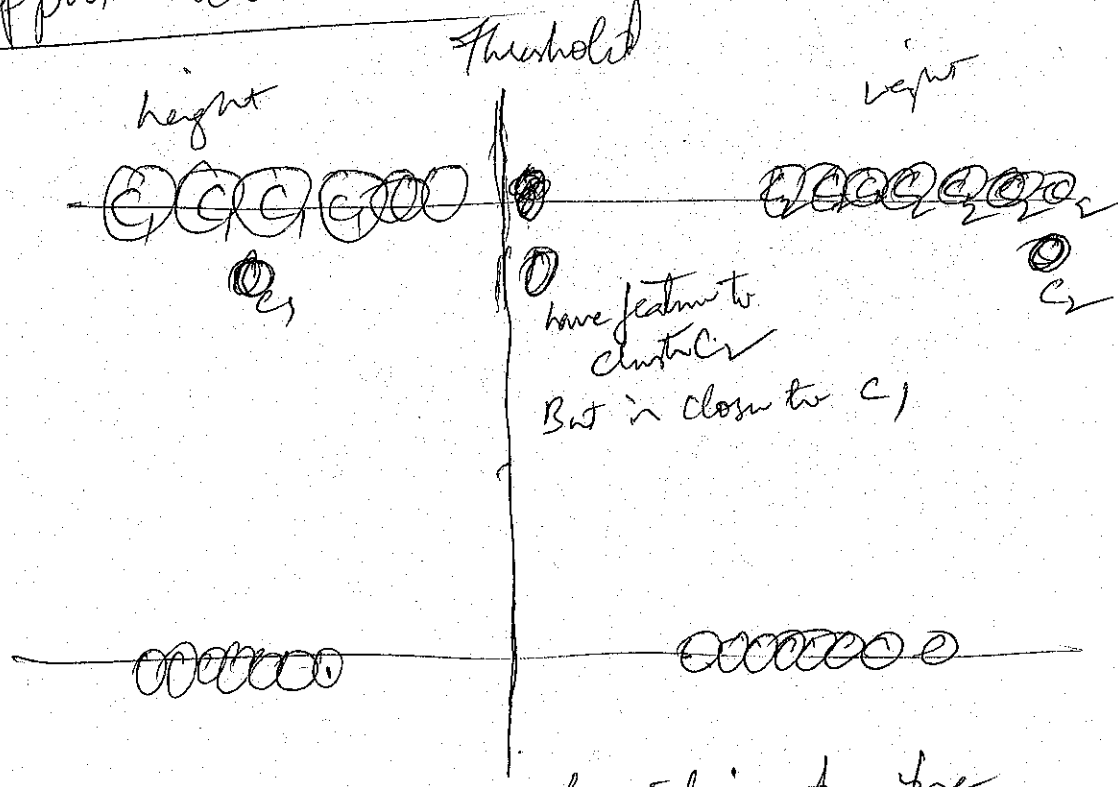
K mediods

	x	y	M_1 distance	M_2	min cost
P_1					
P_2					
P_3					
P_4					
P_5					
P_6					
					Total (cost)

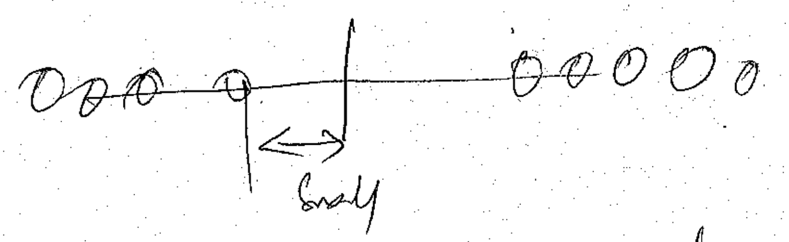
(3)

	docId	words in document	in c=china
Ts	1	Taipei Taiwan	yes
	2	macao Taiwan Shanghai	yes
	3	Japan sapporo	no
	4	sapporo Osaka Taiwan	no
testset	5	Taiwan Taiwan sapporo	?

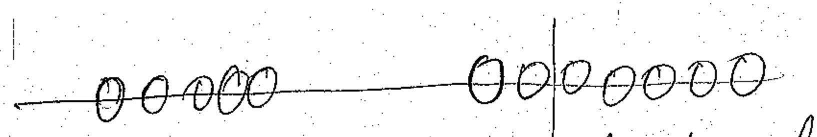
Support Vector Machine



The shortest distance between the observation and the threshold is called the margin.



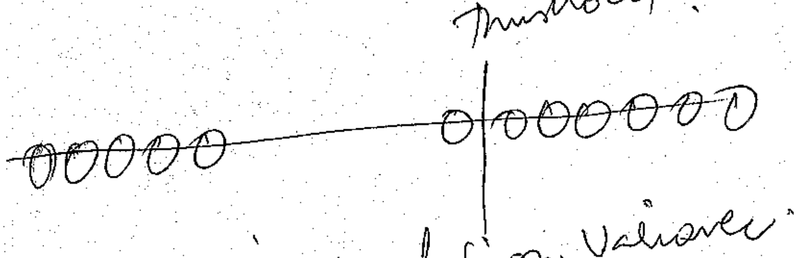
When we use the threshold that gives us the largest margin to make classifications.
we are using a maximal margin classifier.



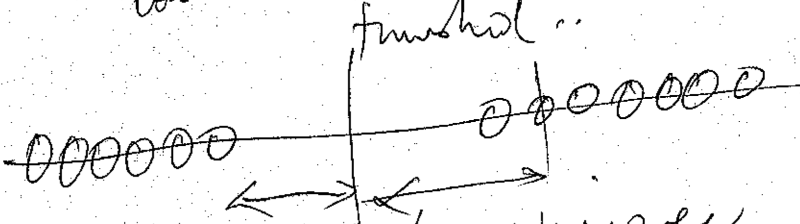
maximal margin classifier are super sensitive to outlier in the training data and that makes them pretty lame.

allow misclassification.

choosing a threshold that allows misclassification is an example of Bias/Variance Tradeoff that plagues all of machine learning.



low bias and high variance.



high bias and low Variance

When we allow misclassification, the distance between the observation and the threshold is called soft margin.

How do we know that this Soft margin is better than this Soft margin

Answer: We use cross validation to determine how many misclassifications and observation to allow inside of soft margin to get the best classification

3-d - plane

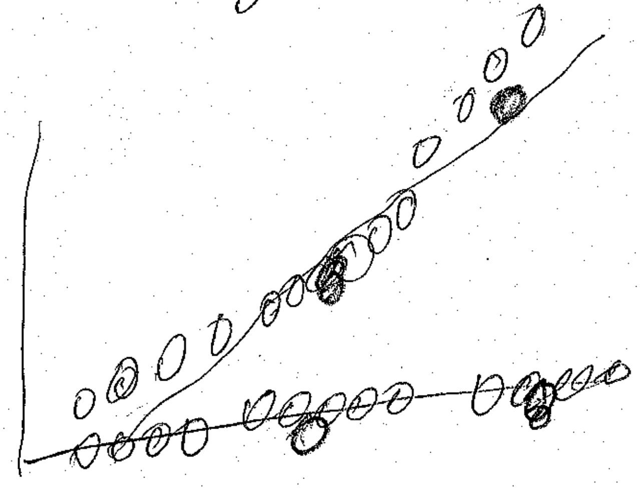
4-d - hyperplane.

Support Vector classifier seem pretty cool because they can handle ~~outlier~~ and because they allow misclassification they can handle overlapping classification.

~~0000 00000 0000~~

Drug in just right

Support vector machines



Dosage = $Dosage^2$

Why not $Dosage^3$

Why not $\sqrt{Dosage}$

In order to make the mathematics possible, support vector machines use something called kernel functions to systematically find support vector classifier in higher dimension.

Polynomial Kernel

When d = 1, the polynomial kernel computes the relationships between each pair of observations in 1-Dimension ~~and~~ these relationships are used for support vector classifier.

d = 2
$dosage^2$ SV classifier

d = 3
dosage SV classifier

d = 4

good value of d is found from cross validation

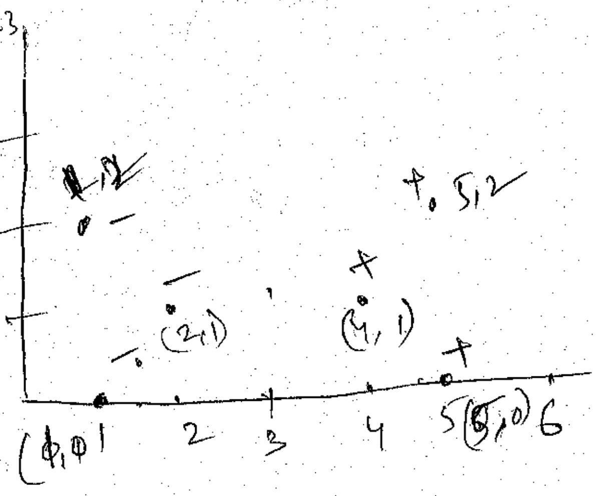
Radial Kernel (Radial Basis function) kernel.

SVM

$-3.5 + 3 \times 0.75 + 0.75 \times 3$

$2.25 + 2.25 = 4.5$

$-3.5 + 0.75 +$



$S_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$S_2 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$

$P_1\ (1, 2)\ -$

$P_2\ (0, 0)\ -$

$P_3\ (2, 1)\ -$

$P_4\ (4, 1)\ +$

$P_5\ (5, 0)\ +$

$P_6\ (5, 2)\ +$

Bias

$\tilde{S}_1 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$

$\tilde{S}_2 = \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix}$

$\alpha_1 \tilde{S}_1 \cdot \tilde{S}_1 + \alpha_2 \tilde{S}_2 \cdot \tilde{S}_1 = -1$

$\alpha_1 \tilde{S}_1 \cdot \tilde{S}_2 + \alpha \tilde{S}_2 \cdot \tilde{S}_2 = +1$

$\alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = -1$

$\alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix} = +1$

$\alpha_1 (4+1+1) + \alpha_2 (8+1+1) = -1$

$\alpha_1 (8+1+1) + \alpha_2 (16+1+1) = +1$

$6\alpha_1 + 10\alpha_2 = -1$

$10\alpha_1 + 18\alpha_2 = +1$

$\cancel{10} - 100\alpha_2 + 108\alpha_2 = 16$

$8\alpha_2 = 16$

$\alpha_2 = \frac{16}{8} = 2$

$\alpha_2 = 2$

$\alpha_1 = \frac{-1 - 10\alpha_2}{6}$

$\frac{-1-20}{6} = -\frac{21}{6}$

$\boxed{\begin{matrix} \alpha_1 = -3.5 \\ \alpha_2 = 2 \end{matrix}}$

$10\left(\frac{-1 - 10\alpha_2}{6}\right) + 18\alpha_2 = 1$

$\frac{-10 - 100\alpha_2}{6} + 108\alpha_2 =$

$\tilde{w} = -3.5\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix}$

$= \cancel{-7\ {-3.5}\ {-3.5} + 8 + 2 + 2}$

$= \begin{pmatrix} -7 \\ -3.5 \\ -3.5 \end{pmatrix} + \begin{pmatrix} 8 \\ 2 \\ 2 \end{pmatrix}$

$= \begin{pmatrix} 1 \\ -1.5 \\ -1.5 \end{pmatrix}$

$\begin{matrix} 1 \\ 1.5 \\ -1.5 \end{matrix}$