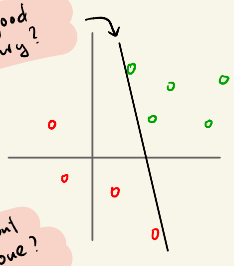
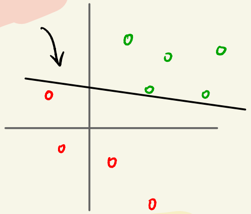
This a good Boundary?

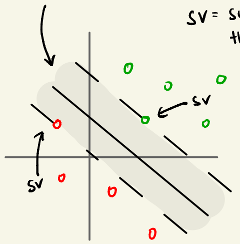


How about this one?



Wait: fit a street between the classes that is as wide as possible.

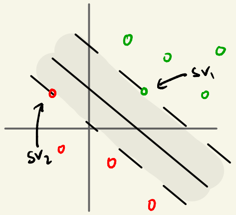
SV = support vector they "support" the hyperplane



Question: if we move some points a tiny bit, for which does the Decision Boundary change? & which not?

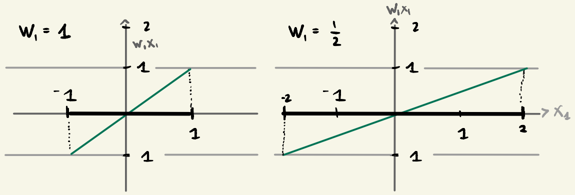
How?

Choose w such that $w^Tx-b=1$ goes through sv_1 and $w^Tx-b=-1$ goes through sv_2.

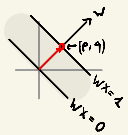


Maximum Margin: maximize the distance from sv_1, sv_2 to the Hyperplane ⇒ minimize $\|w\|$ under conditions $y_i(w^Tx_i-b) \geq 1$

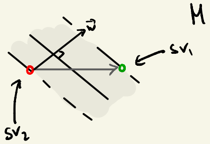
Intuitive picture smaller w ⇒ wider margin



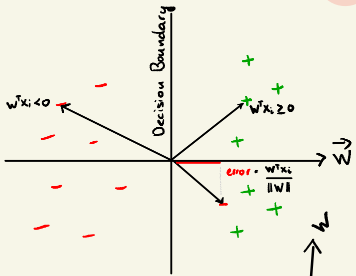
Simple case hyperplane through origin. Below more general.

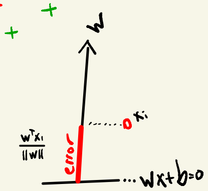


Let $sv_1 = (p,q)$ then $sv_1 = t\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ because $wx=1$ ⇒ $t \cdot w^Tw = 1$ ⇒ $t = \frac{1}{\|w\|^2}$

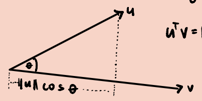


$$\text{Margin} = (sv_1 - sv_2)^T \frac{w}{\|w\|} = sv_1^T\frac{w}{\|w\|} - sv_2^T\frac{w}{\|w\|} = \frac{b+1}{\|w\|} - \frac{b-1}{\|w\|} = \frac{2}{\|w\|}$$



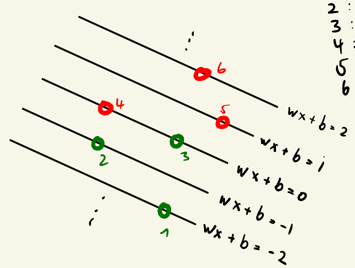


Reminder: Dot Products & Projections



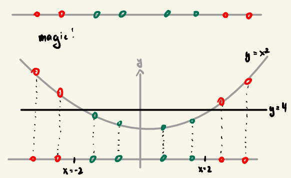
$u^Tv = \|u\|\|v\|\cos\theta$

Errors:
1 : 2
2 : 1
3 : 0
4 : 0
5 : 1
6 : 2



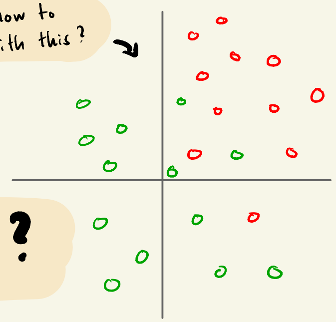
TLDR:

not separable by line   Kernel Trick



Problem! ← impossible if classes are not linearly separable

So..., how to deal with this?

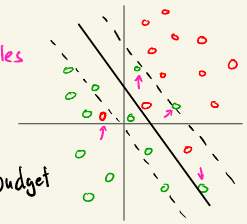


How?

IDEA!

Cut classifier some slack and give it a budet for margin violations and misclassifications

Slack variables ξ_i



$\sum_{i=1} \xi_i \leq$ budget

Hinge Loss

Soft Margin:

minimize:

$\frac{1}{n}C\sum(\max(0, 1-y_i(w^Tx_i-b)) + \|w\|^2$

Look familiar? →

Before (i.e. Ridge)

minimize $A + \lambda B$ bigger λ more weight on B

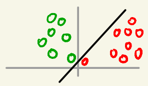
now

minimize $CA + B$ smaller C more weight on B

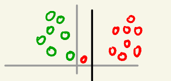
⇒ $C \approx \frac{1}{\lambda}$

Bonus!

also less sensitive to outliers!



⇒ big C → small budget



small C ⇒ bigger budget ⇒ wider margin