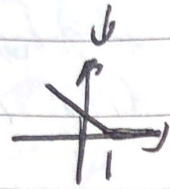
SVM

$$\min_{\theta} C\sum_{i=1}^{m}\left[y^{(i)}cost_1(\theta^T x^{(i)}) + (1-y^{(i)})cost_0(\theta^T x^{(i)})\right] + \frac{1}{2}\sum_{i=1}^{n}\theta_i^2$$



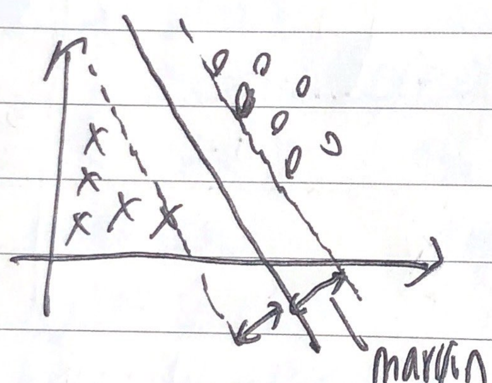
$$\begin{cases} \text{if } y^{(i)}=1, \text{ want } \theta^T x^{(i)} \ge 1 \\ y^{(i)}=0 \qquad \theta^T x^{(i)} \le -1 \end{cases}$$

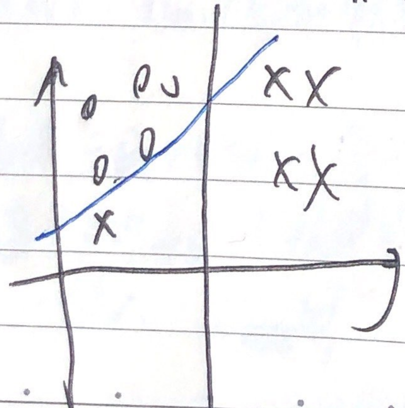
C其实相当于 $\frac{1}{\lambda}$

if C 很大，需要把C乘以的那项=0.

$\min_{\theta} \frac{1}{2}\sum_{i=1}^{n}\theta_i^2$ s.t. $\theta^T x^{(i)} \ge 1$ if $y^{(i)}=1$

$\theta^T x^{(i)} \le 1$ if $y^{(i)}=0$





过拟合

C 很大：有异常点，按蓝线分

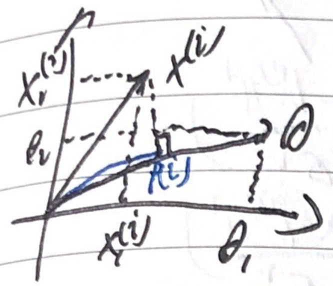
C不是很大：按黑线分（允许异常）

引入松弛变量

设$\theta_0=0$.

$$\min_\theta \frac{1}{2}\sum_{j=1}^{n}\theta_j^2 = \frac{1}{2}(\theta_1^2+\theta_2^2) = \frac{1}{2}\left(\sqrt{\theta_1^2+\theta_2^2}\right)^2 = \frac{1}{2}\|\theta\|^2$$

范数

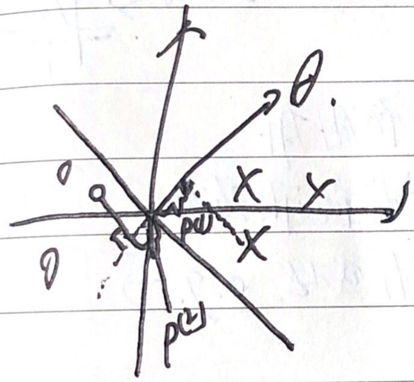


$$\theta^T x^{(i)} = p^{(i)}\cdot\|\theta\|$$
$$= \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)}$$

$$\min_\theta \frac{1}{2}\sum_{j=1}^{n}\theta_j^2 = \frac{1}{2}\|\theta\|^2$$

s.t. $p^{(i)}\cdot\|\theta\| \geq 1$ if $y^{(i)}=1$

$p^{(i)}\cdot\|\theta\| \leq -1$ if $y^{(i)}=0$

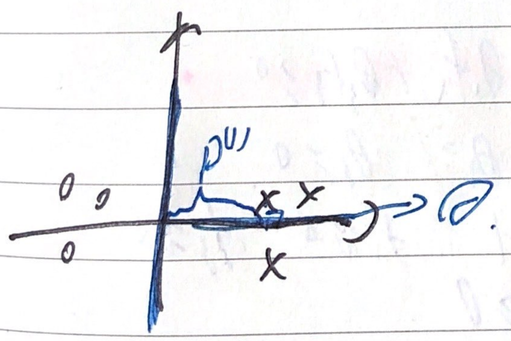


$p^{(i)}\cdot\|\theta\| \geq 1$

$p^{(i)}$很小 → $\|\theta\|$很大

但我们要 $\min \frac{1}{2}\|\theta\|^2$

∴要$p^{(i)}$大.



$p^{(i)}$大

$\|\theta\|$小

kernel

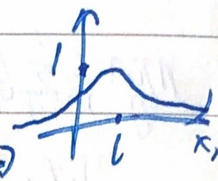
new feature f_1, f_2, f_3

$$\sum_{j=1}^{n}(x_j - l_j^{(1)})^2$$

$$f_1 = \text{similarity}(x, l^{(1)}) = \exp\left(-\frac{\|x-l^{(1)}\|^2}{2\sigma^2}\right)$$

$$f_2 = \text{similarity}(x, l^{(2)}) = \exp\left(-\frac{\|x-l^{(2)}\|^2}{2\sigma^2}\right)$$

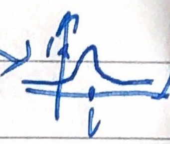
高斯核函数

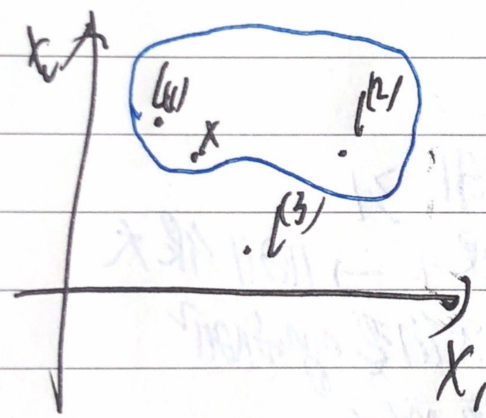


$l^{(1)}, l^{(2)}, l^{(3)}$：标记点

σ很大：feature大 变化慢，

σ：函数值跌落到0的速率. σ很小：过拟合 →

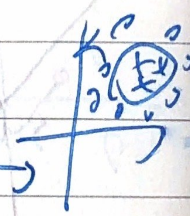




圈内预测为1

这样就可以预测

非线性. e.g. →



predict 1 when $\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$

$\theta_0 = -0.5 \quad \theta_1 = 1 \quad \theta_2 = 1 \quad \theta_3 = 0.$

x接近 $l^{(1)}$, $f_1 \approx 1 \quad f_2 \approx 0 \quad f_3 \approx 0$

$\theta_0 + \theta_1 = 0.5 \geq 0$

x被标记为1.

$l^{(1)}, l^{(2)} \ldots$ 是每一个样本点

given example $(x^{(i)}, y^{(i)})$

$f_1^{(i)} = sim(x^{(i)}, l^{(1)})$

$f_2^{(i)} = sim(x^{(i)}, l^{(2)})$

$\vdots$

$f_m^{(i)} = sim(x^{(i)}, l^{(m)})$ m个样本点

$$\min_\theta C \sum_{i=1}^{m} y^{(i)} cost_1(\underline{\theta^T f^{(i)}}) + (1-y^{(i)}) cost_0(\underline{\theta^T f^{(i)}}) + \frac{1}{2}\sum_{j=1}^{n=m}\theta_j^2$$

过拟合时：减小 C，增大 σ^2

n = number of features

m = number of training example

① $n \geq m$：用 logistic regression

② n is small, m is intermediate：用 SVM with 高斯核函数

$n = 1-1000$, $m = 10-10000$

③ n is small, m is large：$n = 1-1000$, $m = 50000+$

create more features

then use logistic regression or SVM without kernel

∵带核函数样本太大运行慢