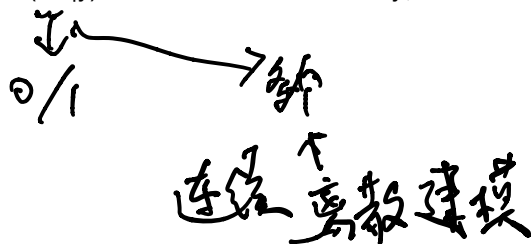


回顾上一讲的 垃圾邮件分类，首先是把原来邮件的 词出现 做成列向量，再假设两两独立， $P(X|Y)=\text{multi } P(X_{ij}|y)$ ，然后通过极大似然 去找 y ，



每个词指向字典中的 index

原来这种0/1的 多元伯努利事件模型
有多种取值的 多项式事件模型

$$P(x, y) = \left(\prod_{i=1}^n P(x_i | y) \right) P(y)$$

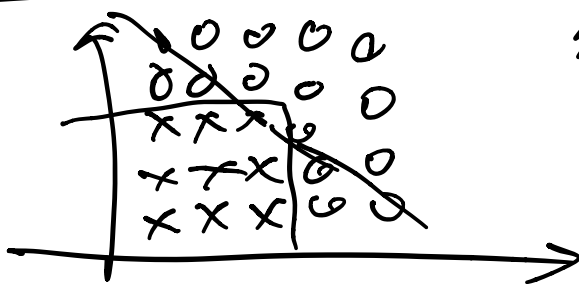
$$\phi_{k|y=1} = P(x_j = k | y=1)$$

$$\phi_{k|y=0} = P(x_j = k | y=0)$$

$\phi_{k|y=1}$ 是在垃圾邮件中词出现的次数
 $\phi_{k|y=0}$ 是在非垃圾邮件中词出现的次数
 k 取的是 = 词典中的词数?

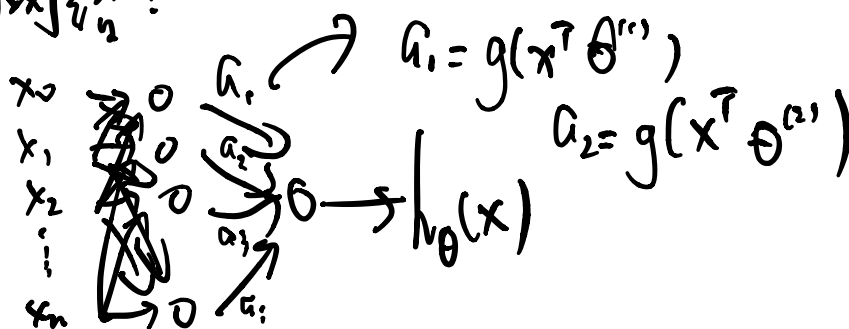
文本分类这样的一元模型（不考虑位子 只考虑 次数 或者 是否出现）从效果上已经很好了

非线性分类器



线性地: $x_0, x_1, x_2, x_3, \dots$ 通过一个映射 $\rightarrow 0 \rightarrow$

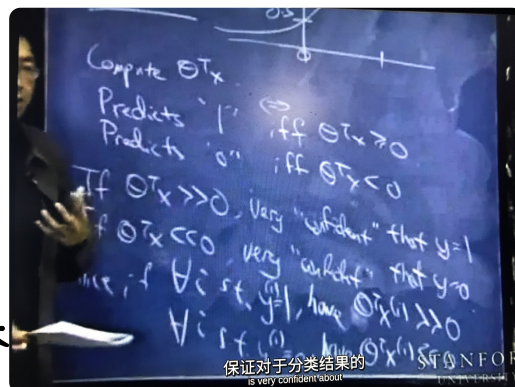
神经网络 h_{θ} :



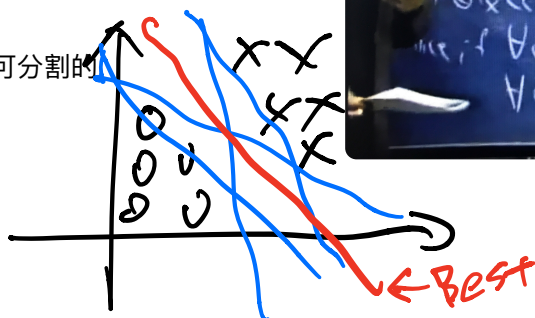
支持向量机发明之前很高效的算法 数字识别，语音学习

支持向量机

如果分类器不只 分类 还能很确定的分类 则是nice的



先假设 训练集合 都是线性可分割的



$$y \in \{-1, +1\}, h(x) \in \{-1, +1\}$$

$$g(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{v.w} \end{cases}$$

$$h_{w,b}(x) = g(w^T x + b)$$

函数间隔(function margin), 一个超平面(w,b) 和特定训练样本 相关的函数间隔

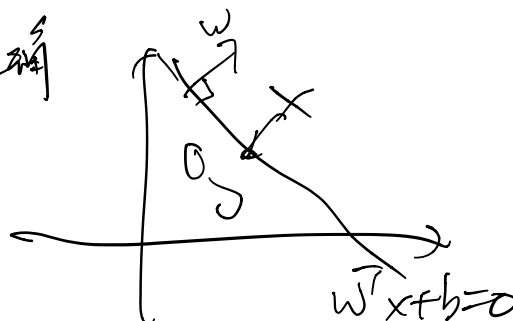
$$\hat{g}^{(i)} = y^{(i)} (w^T x^{(i)} + b)$$

$$\text{if } y^{(i)} = 1 \quad \text{want } w^T x^{(i)} + b \gg 0$$

$$\text{if } y^{(i)} = -1 \quad \text{want } w^T x^{(i)} + b \ll 0$$

$$\text{if } \hat{g}^{(i)} \geq 0 \rightarrow \text{正确}$$

$$\hat{g} = \min_i \hat{g}^{(i)}$$



$$w^T x^{(i)} + b = \int \frac{w^T \omega}{\|w\|} = \int^{(i)} \|w\|$$

几何解释

$$f^{(i)} = \gamma^{(i)} \left[\frac{w^T}{\|w\|} x^{(i)} + \frac{b}{\|w\|} \right]$$

$$\rightarrow f^{(i)} = \frac{\hat{f}^{(i)}}{\|w\|}$$

函数间隔, 由 $\|w\|$ 受

↓
不变