4/20/18

Note Title 1/15/2000

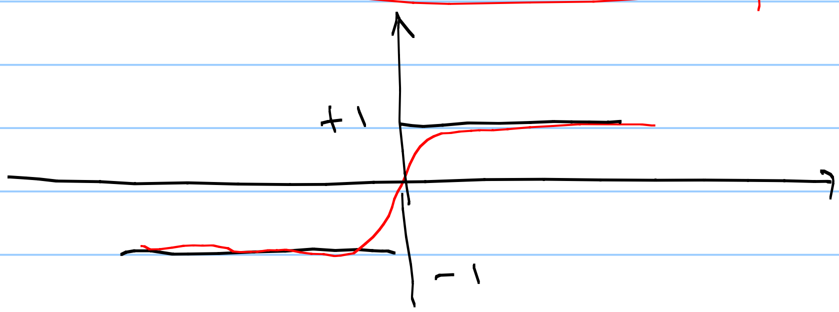
Logistic regression perspective to classification

$$\operatorname{sign}\left(b+\bar{x}_p^T \bar{w}\right) = \begin{cases} 1, & b+\bar{x}_p^T\bar{w} > 0 \\ -1, & b+\bar{x}_p^T\bar{w} < 0 \end{cases}$$

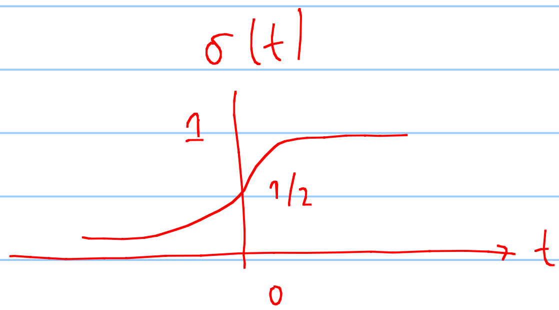
We want to find a boundary so that

$$\operatorname{sign}\left(b+\bar{x}_p^T \bar{w}\right) = y_p , \quad p=1,\dots,P$$

$$\Rightarrow \boxed{\operatorname{sign}\left(y_p\left(b+\bar{x}_p^T\bar{w}\right)\right) = 1}$$

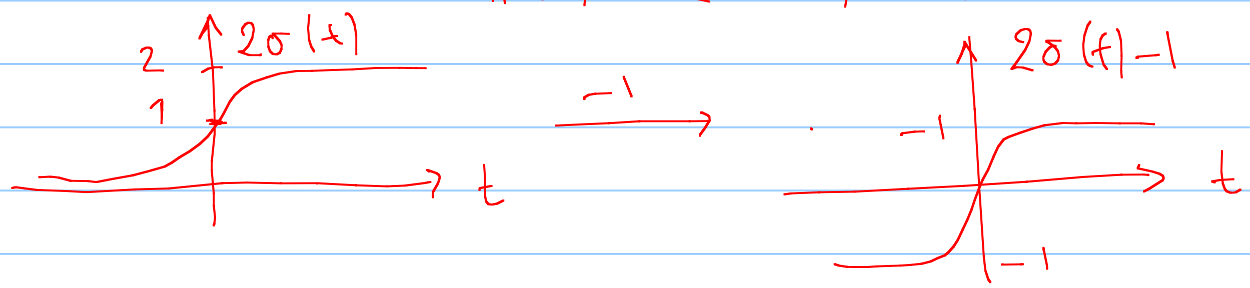


Find smooth approximators



$$\sigma(t) = \frac{1}{1+e^{-t}}$$

$$\tanh(t) = 2\sigma(t) - 1$$



Replace sign$(\cdot)$ by tanh$(\cdot)$

$$\tanh\left(y_p\left(b+\bar{x}_p^T\bar{w}\right)\right) \simeq 1$$

$$2\sigma(\cdot)-1 \simeq 1$$

$$\rightarrow 2\sigma(\cdot) \simeq 2$$

$$\rightarrow \sigma(\cdot) \simeq 1$$

$$\rightarrow \frac{1}{1+e^{-t}} \simeq 1 \Rightarrow 1+e^{-t} \simeq 1$$

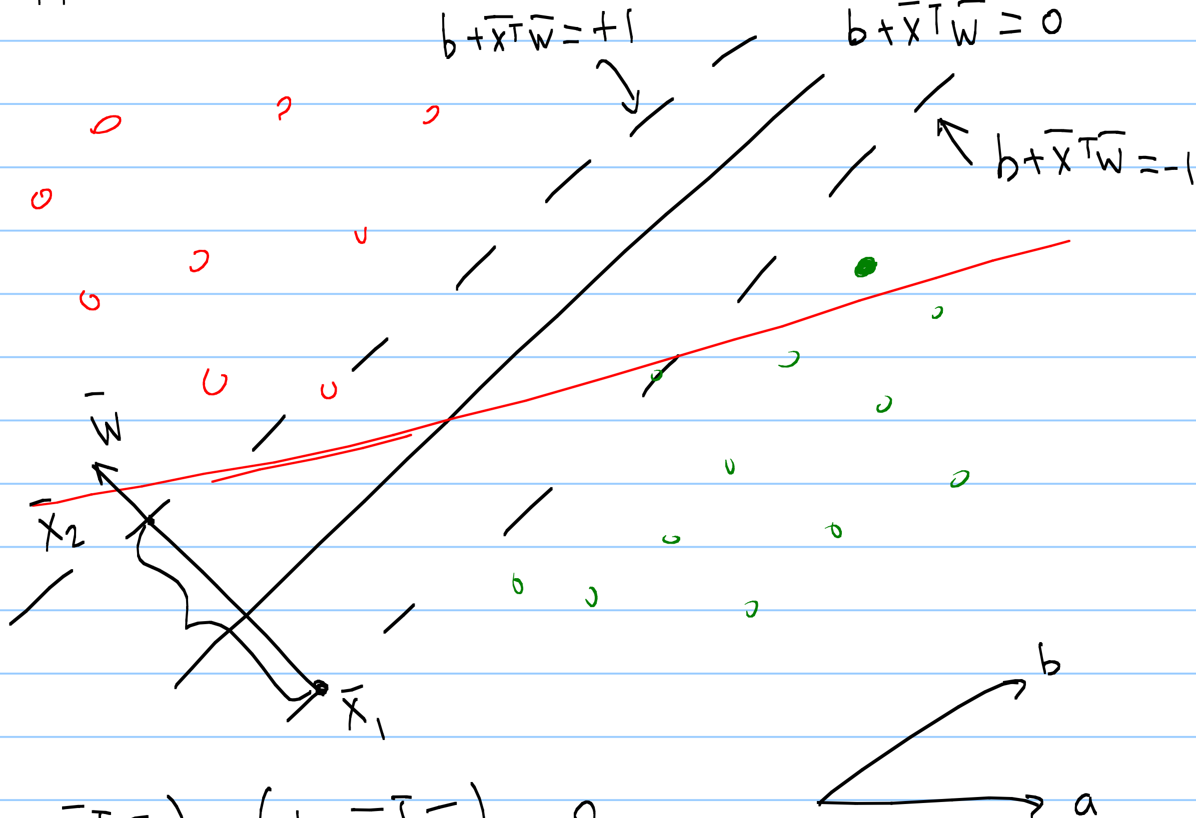
$$\rightarrow \log\left(1+e^{-t}\right) \simeq 0$$

$$\therefore \log\left(1+e^{-y_p\left(b+\bar{x}_p^T\bar{w}\right)}\right) = 0 \quad \forall_p$$

$$\therefore g(b,\bar{w}) = \sum_{p=1}^{P} \log\left(1+e^{-y_p\left(b+\bar{x}_p^T\bar{w}\right)}\right)$$

softmax classifier

Support Vector Machine Classifier



$$\left(b + \bar{x}_2^T \bar{w}\right) - \left(b + \bar{x}_1^T \bar{w}\right) = 2$$

$$\Rightarrow \left(\bar{x}_2 - \bar{x}_1\right)^T \bar{w} = 2$$

$$\Rightarrow \|\bar{x}_1 - \bar{x}_2\| \cdot \|w\| \cdot \overset{1}{\cancel{\cos 0^\circ}} = 2$$

$$\Rightarrow \|\bar{x}_1 - \bar{x}_2\| = \frac{2}{\|w\|}$$

$\Big(\bar{a}^T \bar{b} = \|a\|_2 \cdot \|b\|_2 \cdot \cos\theta$

∴ In order to find the hyperplane w/ the largest margin we need to find one that separates the data w/ the smallest length $\bar{w}$

Mathematically

hard margin SVM

$$\boxed{\begin{aligned} &\min_{b,\bar{w}} \|\bar{w}\|_2^2 \\ &\text{s.t.} \quad \max\left(0,\, 1 - y_p\left(b + \bar{x}_p^T \bar{w}\right)\right) = 0 \end{aligned}}$$

Relaxed form of soft-margin SVM

$$g(b,\bar{w}) = \sum_{p=1}^{P} \max\left(0,\, 1 - y_p\left(b + \bar{x}_p^T \bar{w}\right)\right) + \lambda \|w\|_2^2$$

Soft-max approximation (soft-margin SVM)

(log-loss SVM)

$$\underbrace{\sum \log\left(1 + e^{-y_p\left(b + \bar{x}_p^T \bar{w}\right)}\right)}_{\substack{\text{soft-max classifier} \\ \text{logistic regression classifier}}} + \lambda \|w\|_2^2$$