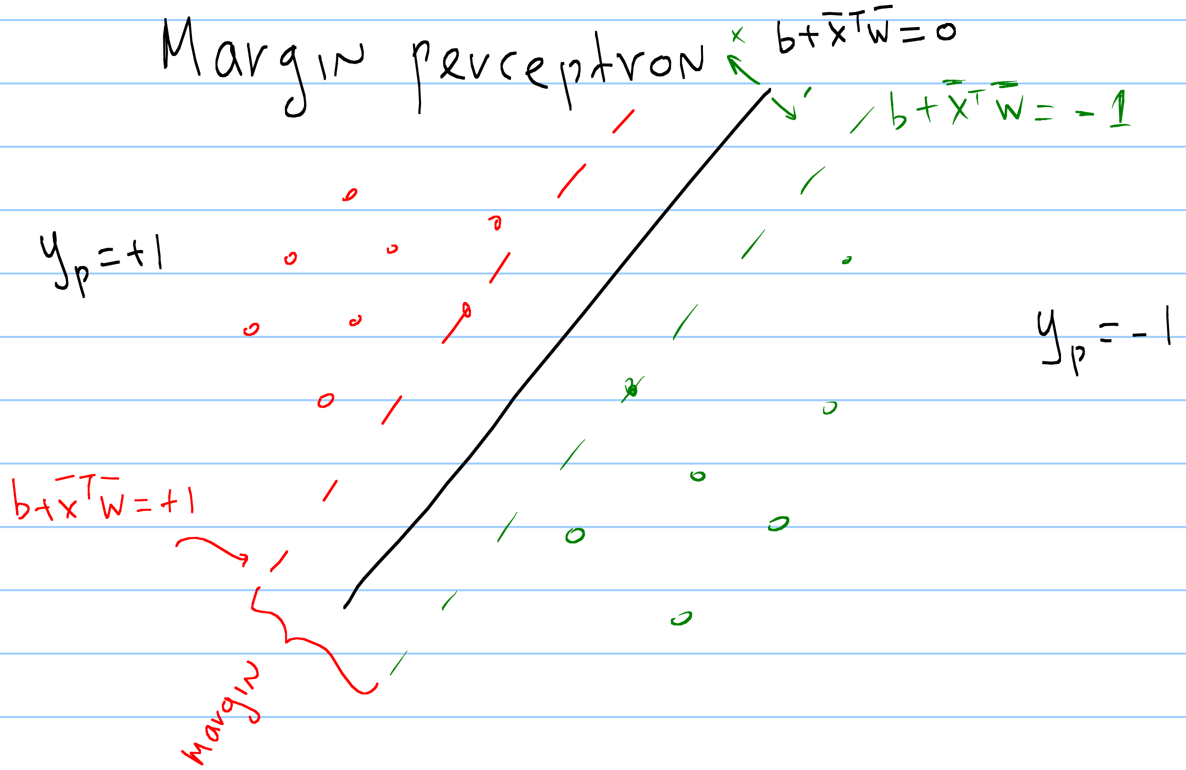
Note Title 4/18/18 1/13/2000

Margin perceptron



given $\{(\bar{x}_p, y_p)\}_{p=1}^{P}$

find $b, \bar{w}$ s.t.

$$b+\bar{x}_p^T\bar{w} \geq 1 \quad \text{if } y_p=+1$$

$$b+\bar{x}_p^T\bar{w} \leq -1 \quad \text{if } y_p=-1$$

$$\Rightarrow y_p(b+\bar{x}_p^T\bar{w}) \geq 1 \longrightarrow 1-y_p(b+\bar{x}^T\bar{w}) \leq 0$$

$$\Rightarrow \max(0, 1-y_p(b+\bar{x}_p^T\bar{w}))=0$$

Cost function for margin perceptron (hinge cost)

$$g_3(b,\bar{w}) = \sum_{p=1}^{P} \max(0, 1-y_p(b+\bar{x}_p^T\bar{w}))$$

$$\therefore b^*, \bar{w}^* = \arg\min_{b,\bar{w}} g_3(b,\bar{w})$$

2 approximations to margin perceptron

1) $\text{soft}(s_1, s_2) \approx \max(s_1, s_2)$

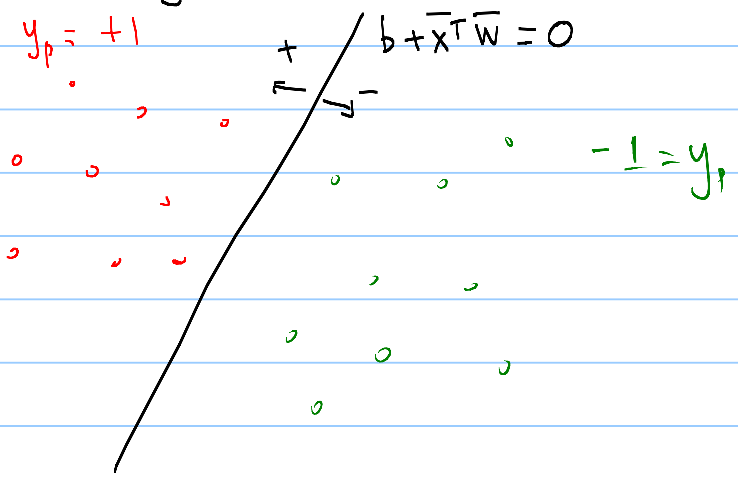
$$\max\left(0,\ 1 - y_p\left(b + \bar{x}_p^T \bar{w}\right)\right) \approx \text{soft}\left(0,\ 1 - y_p\left(b + \bar{x}^T \bar{w}\right)\right)$$

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

2) Square margin perceptron cost

$$g_4(b, \bar{w}) = \sum_{p=1}^{P} \max^2\left(0,\ 1 - y_p\left(b + \bar{x}_p^T \bar{w}\right)\right)$$

The accuracy of a learned classifier



Correct classification

$$\text{sign}\left(b + \bar{x}_p^T \bar{w}\right) = y_p$$

if incorrect classification

$$\text{sign}\left(b + \bar{x}_p^T \bar{w}\right) = -y_p$$

combine

$$\rightarrow \quad \operatorname{sign}\left(-y_p\left(b+\bar{x}_p^T\bar{w}\right)\right) = \begin{cases} +1 & \text{if } \bar{x}_p \text{ is classified incorrectly} \\ -1 & \text{if } \bar{x}_p \text{ is classified correctly} \end{cases}$$

To count number of misclassifications

$$g_5(b,\bar{w}) = \sum_{p=1}^{P} \max\left(0,\ \operatorname{sign}\left(-y_p\left(b+\bar{x}_p^T\bar{w}\right)\right)\right)$$

$$0 \le g_5(b,\bar{w}) \le P$$

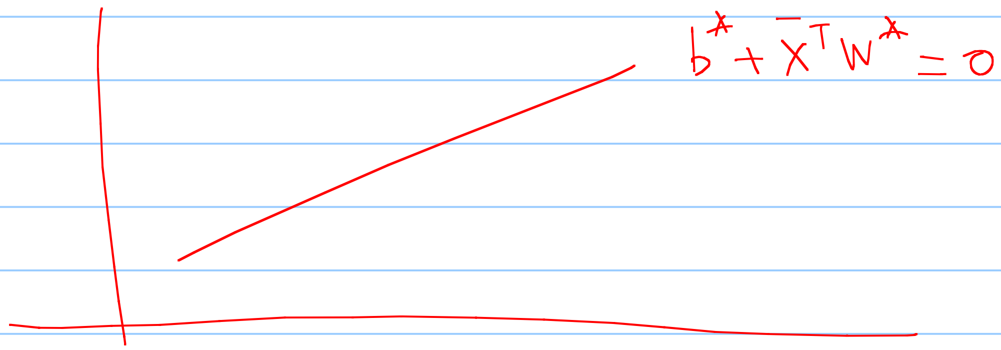
Use g_5 to define parameters of classifier

$$\min_{b,\bar{w}} g_5(b,\bar{w})$$ → non-convex → highly discontinuous

Accuracy of classifier

$$\text{accuracy} = 1 - \frac{g_5(b^*,\bar{w}^*)}{P}$$

$$0 \le \text{accuracy} \le 1$$



$$y_{new} = \operatorname{sign}\left(b^* + \bar{x}_{new}^T w^*\right)$$