4/27/20

2-class classification

- regression perspective (expanded view)
- perceptron perspective ("top-down" view of feature space)

A: {0, 1} label

A.1 regression perspective

logistic regression → using the LS cost

$$g(\tilde{w}) = \frac{1}{P}\sum_{p=1}^{P}\left(\sigma(\tilde{x}_p^T\tilde{w}) - y_p\right)^2$$

non-convex, OK

log-error
cross-entropy
cost

$$\boxed{g(\tilde{w}) = \frac{-1}{P}\sum_{p=1}^{P} y_p \left(\log\left(\sigma(\tilde{x}_p^T\tilde{w})\right)\right) + (1-y_p)\log\left(1-\sigma(\tilde{x}_p^T\tilde{w})\right)}$$

convex

B. $\underline{\{-1, +1\}}$ labels

B.1. regression perspective

- $\text{sign}(\cdot) \to \tanh(\cdot)$

- log error

$$\boxed{\begin{array}{l}\text{Soft-max logistic regression} \\ g(\tilde{w}) = \frac{1}{P}\sum_{p=1}^{P} \log\left(1+e^{-y_p \tilde{x}_p^T \tilde{w}}\right)\end{array}}$$

convex

$\underline{\text{B.2}}$ perceptron perspective

"Issue" w/ soft-max logistic regression

magnitude of weights $\rightarrow \infty$

why:

assume that data are perfectly separable

$\Rightarrow \boxed{\tilde{w}^*} : \boxed{\tilde{x}_p^T \tilde{w}^* = 0} \Leftarrow$

clearly: $\tilde{w}^* \rightarrow c\,\tilde{w}^* : c\,\tilde{x}_p^T \tilde{w} = 0$

$\boxed{c > 1}$

LR: $e^{-y_p \tilde{x}_p^T \tilde{w}^*} > e^{-c\, y_p \tilde{x}_p^T \tilde{w}^*}$

$$e^{-10} > e^{-20}$$

$$\lim_{c \to \infty} e^{-c\, y_p \tilde{x}_p \tilde{w}^*} = 0$$

$e^{-y_p \underbrace{\tilde{x}_p^T \tilde{w}^*}_{0}} = e^{-0} = 1$

$$\Rightarrow g(\tilde{w}) = \frac{1}{P} \sum_{p=1}^{P} \log\left(1 + e^{-y_p \tilde{x}_p^T \tilde{w}^*}\right)$$

0

$\log(1+1) = \log 2$

$\tilde{w}^0 = \tilde{w}^*$ $\log 1 = 0$

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

- for $\tilde{x}_p^T \tilde{w}^* = 0$, for all p

minimum of $g(\tilde{w}) = \frac{1}{P} \sum_{p=1}^{P} \log(1+1) = \underline{\log 2}$

- However, for $\tilde{w}^* \longrightarrow C\tilde{w}^*$, $C > 1$

$$\lim_{C \to \infty} e^{-C y_p \tilde{x}_p^T \tilde{w}} = 0$$

minimum of $g(\tilde{w}) = \frac{1}{P} \sum_{p=1}^{P} \log(1+0) = \underline{0}$

$\Rightarrow$ the algorithm converges to $\|\tilde{w}\| \to \infty$

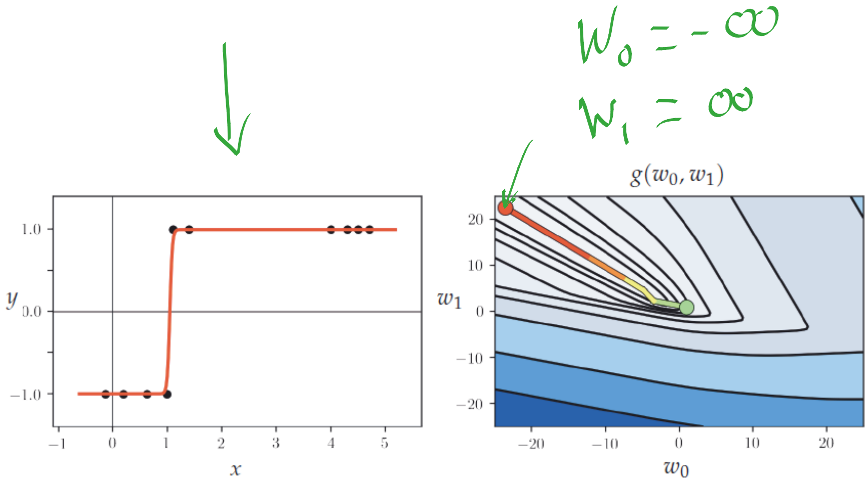


Figure 6.13 (top row) Figure associated with Example 6.6. See text for details.

Address the issue

Through regularization

- early stopping criterion
- use of regularizers

constrained optimization

$$\begin{cases} \min_{\tilde{w}} \dfrac{1}{P} \sum_{p=1}^{P} \log\left(1 + e^{-y_p\left(\underline{b} + \bar{x}_p^T \underline{\bar{w}}\right)}\right) \\ \text{s.t. } \underline{\|\bar{w}\|_2 = \epsilon} \end{cases}$$

Relax it

hyperparameter

$$\min_{\tilde{w}} \left\{ \underbrace{\frac{1}{P} \sum_{p} \log(\cdot)} + \lambda \underbrace{\left(\|\bar{w}\|_2 - \epsilon\right)^2} \right\}$$

regularization parameter

or

Lagrange multiplier

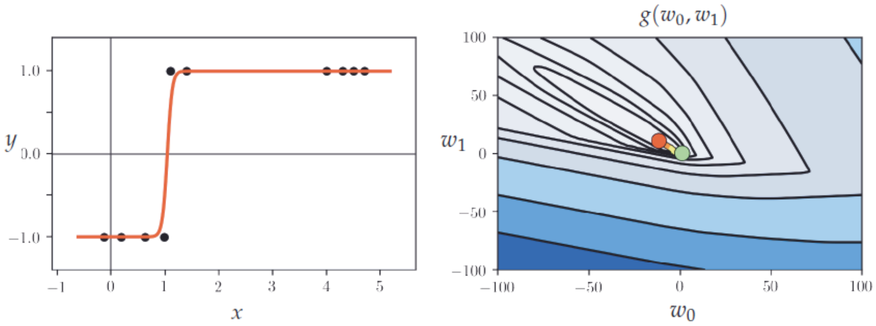
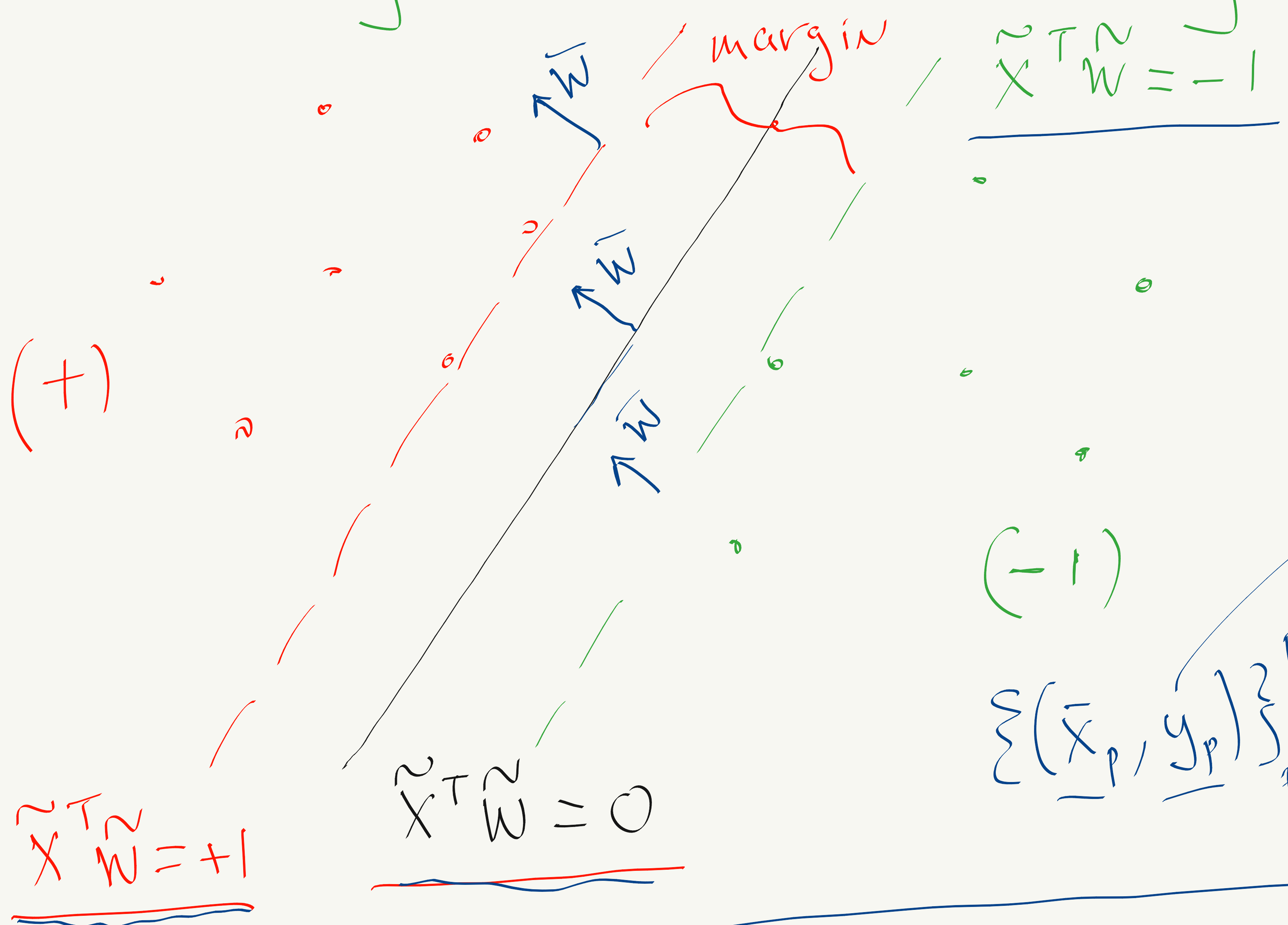


Figure 6.16 Figure associated with Example 6.7. See text for details.

Support Vector Machine (SVM) classifier

Margin Perceptron Classifier

$\Rightarrow$ assuming classes are linearly separable



$$\tilde{w} = \begin{bmatrix} b \\ \bar{w} \end{bmatrix}$$

$y_p \in \{-1, 1\}$

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

$\tilde{x}^T\tilde{w} = \varepsilon$

$\Rightarrow \tilde{x}^T\left(\frac{\tilde{w}}{\varepsilon}\right) = 1$

$\hookrightarrow \tilde{w}'$

$\therefore$

$$\boxed{\begin{array}{lll} \tilde{x}_p^T\tilde{w} \geq 1 & \text{assign} & y_p = +1 \\ \tilde{x}_p^T\tilde{w} \leq -1 & \text{"} & y_p = -1 \end{array}}$$

$$y_p\left(\tilde{x}_p^T\tilde{w}\right) \geq 1 \Rightarrow 1 - y_p\left(\tilde{x}_p^T\tilde{w}\right) \leq 0$$

$$\Rightarrow \max\left(0, 1 - y_p\left(\tilde{x}_p^T\tilde{w}\right)\right) = 0$$

$$\boxed{\text{Cost function for margin perceptron} \quad g(\tilde{w}) = \left(\frac{1}{P}\right)\sum_{p=1}^{P} \max\left(0, \underbrace{1}_{\epsilon} - y_p \tilde{x}_p^T \tilde{w}\right)}$$

$\epsilon = 0$, perceptron

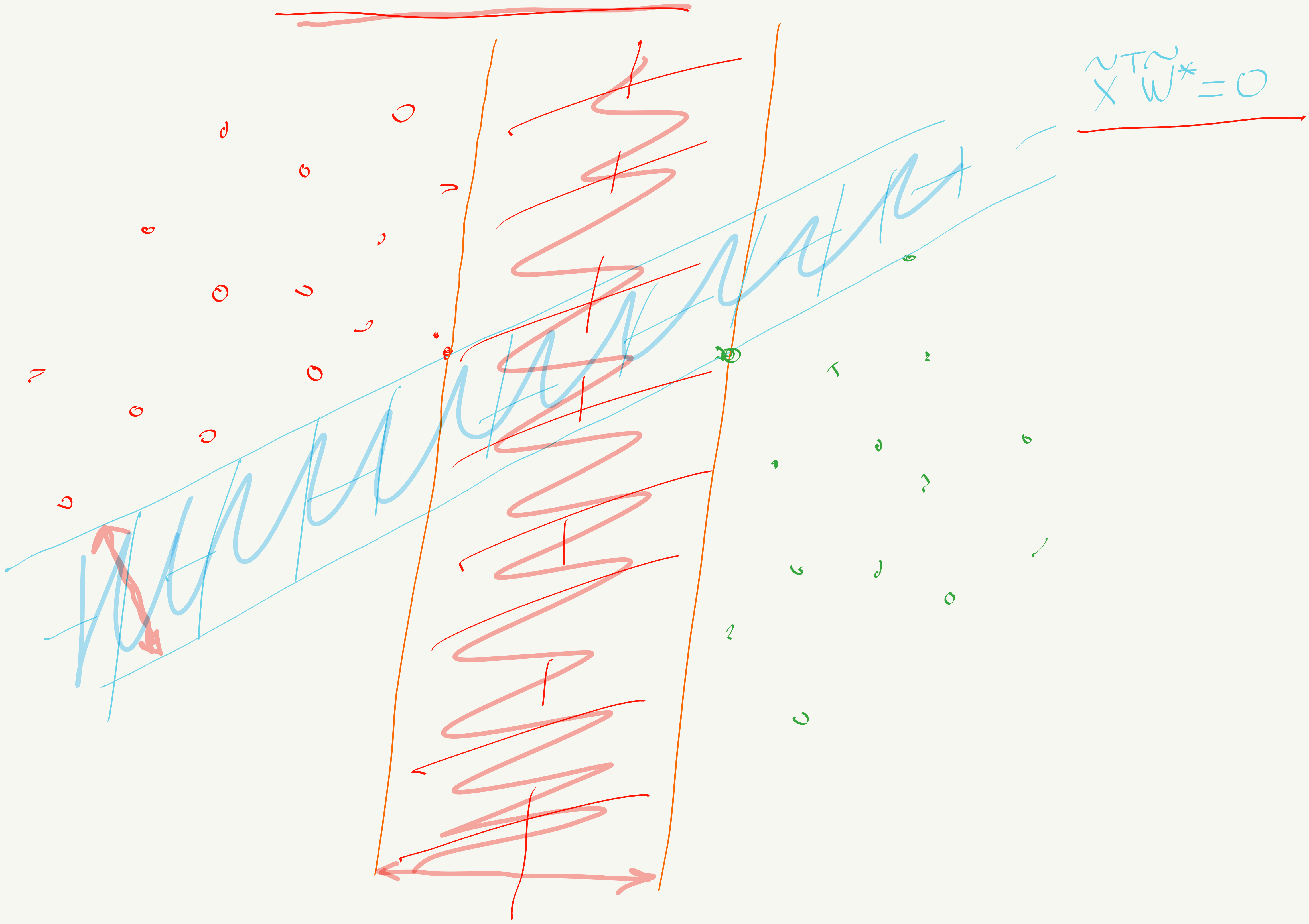
Approximations

- $\text{soft}(s_1, s_2) \simeq \max(s_1, s_2)$

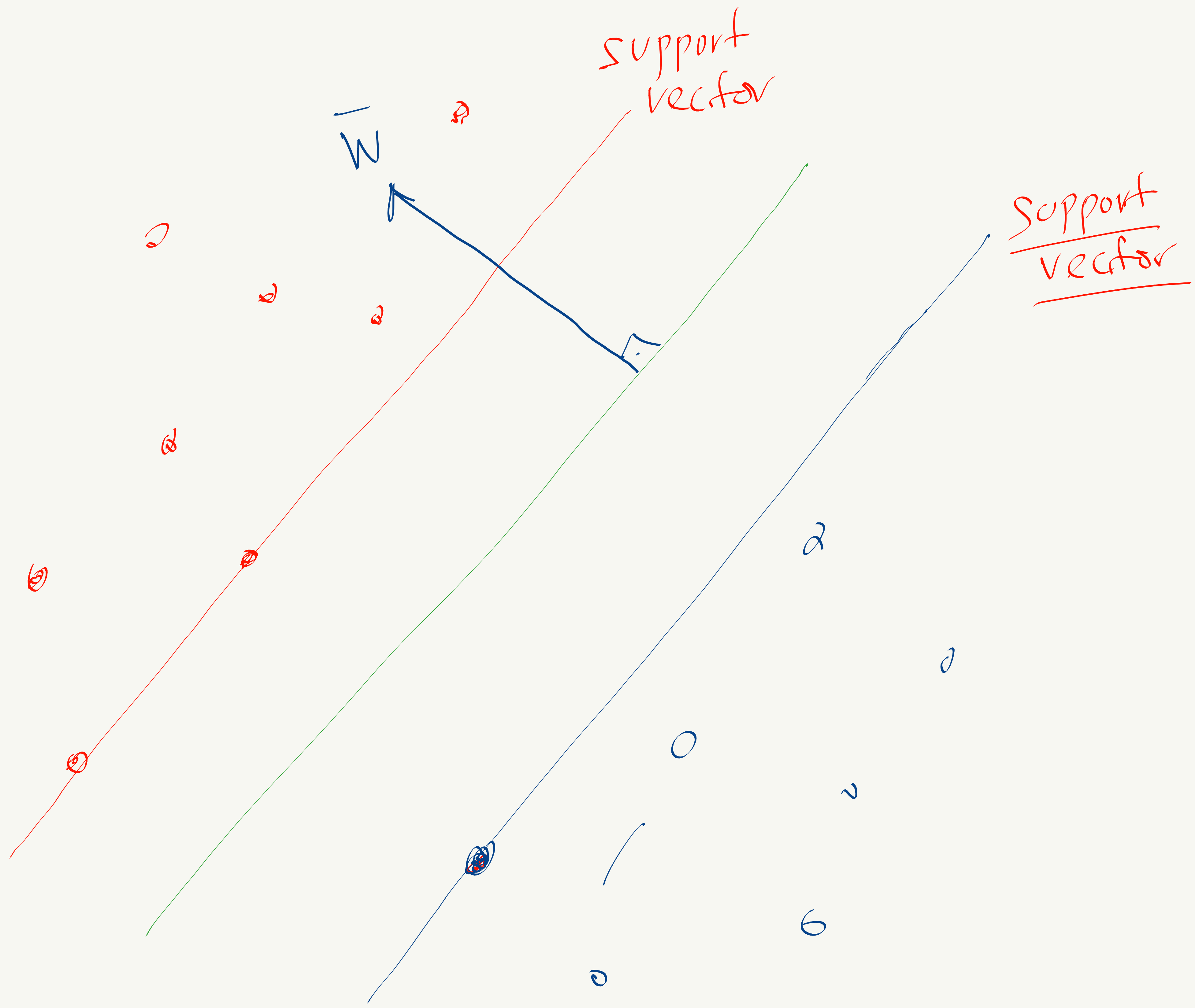
$$\Rightarrow \max(0, 1 - y_p \tilde{x}_p^T \tilde{w}) = \log\left(1 + e^{1 - y_p \tilde{x}_p^T \tilde{w}}\right)$$

$$\Rightarrow \boxed{g(\tilde{w}) = \frac{1}{P}\sum_{p=1}^{P} \log\left(1 + e^{1 - y_p \tilde{x}_p^T \tilde{w}}\right)}$$

SVM



$\tilde{x}^T \tilde{w}^{**} = 0$

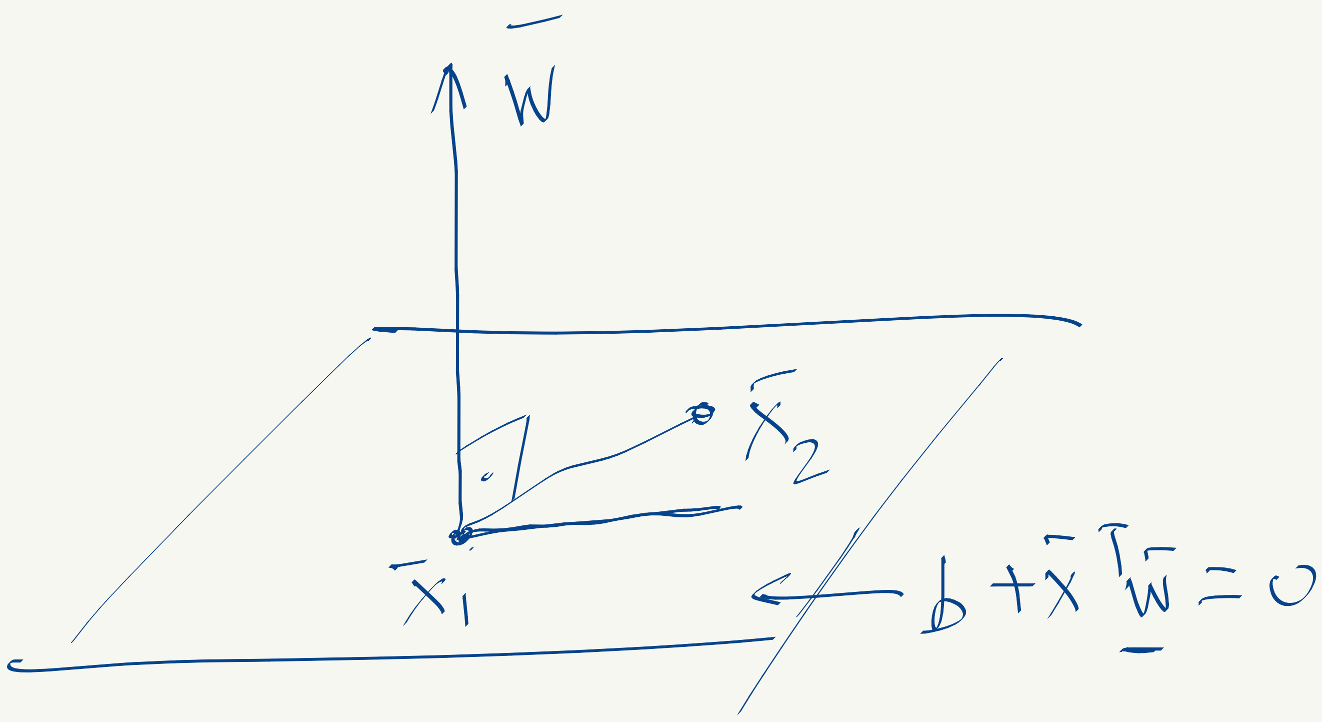


$$\underline{b + \bar{x}^T \bar{w} = 0}$$

$$\equiv$$

$$\tilde{x}^T \tilde{w} = 0$$

$$\tilde{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_N \end{bmatrix} \qquad \tilde{w} = \begin{bmatrix} b \\ w_1 \\ \vdots \\ w_N \end{bmatrix}$$



$\bar{x}_1 \in$ hyperplane : $b + \bar{x}_1^T \bar{w} = 0$

$\bar{x}_2 \in$ " : $b + \bar{x}_2^T \bar{w} = 0$

$$\left.\begin{array}{l} b + \bar{x}_1^T \bar{w} = 0 \\ b + \bar{x}_2^T \bar{w} = 0 \end{array}\right\} \ \underline{(\bar{x}_1 - \bar{x}_2)^T \bar{w}} = 0$$