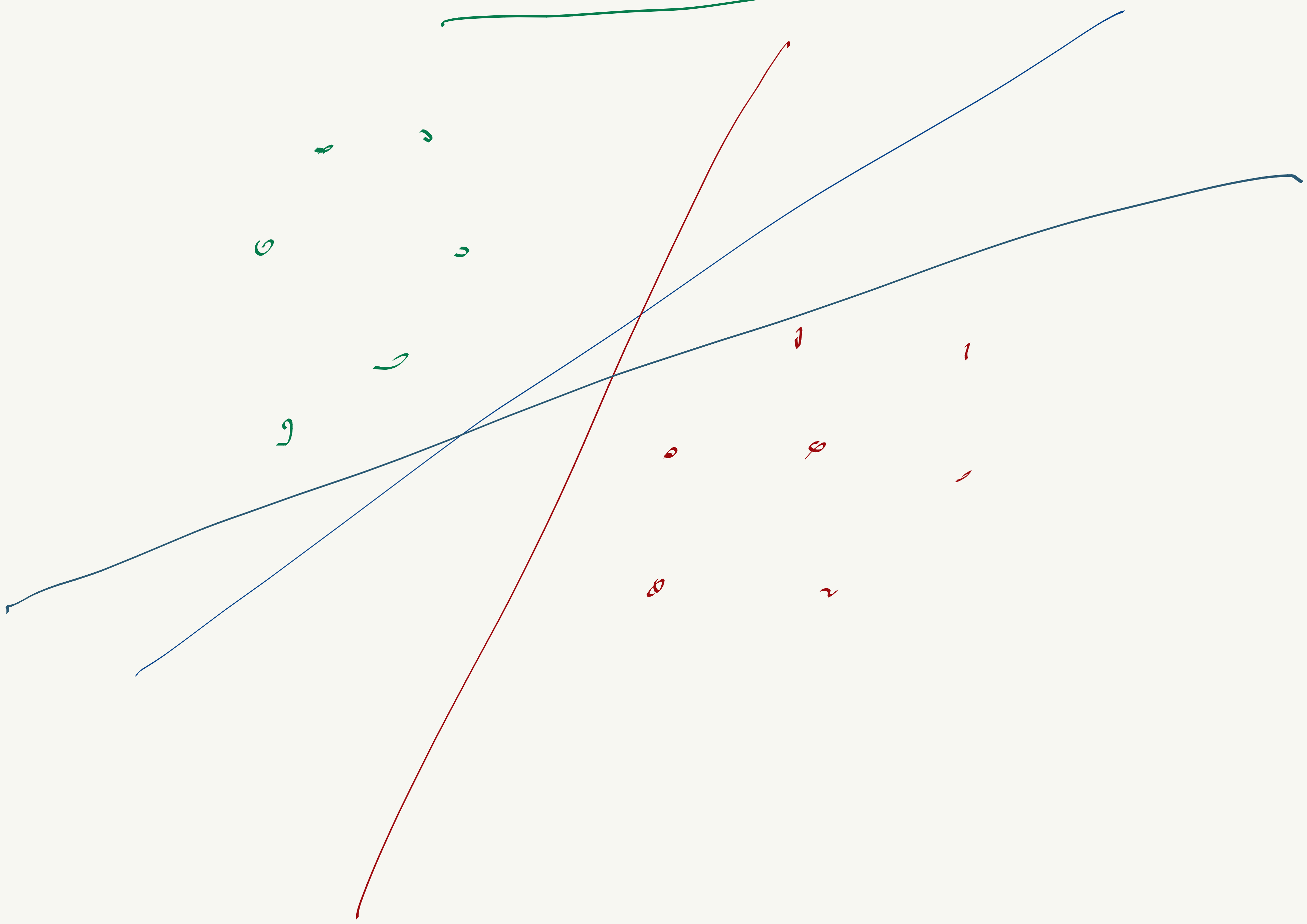
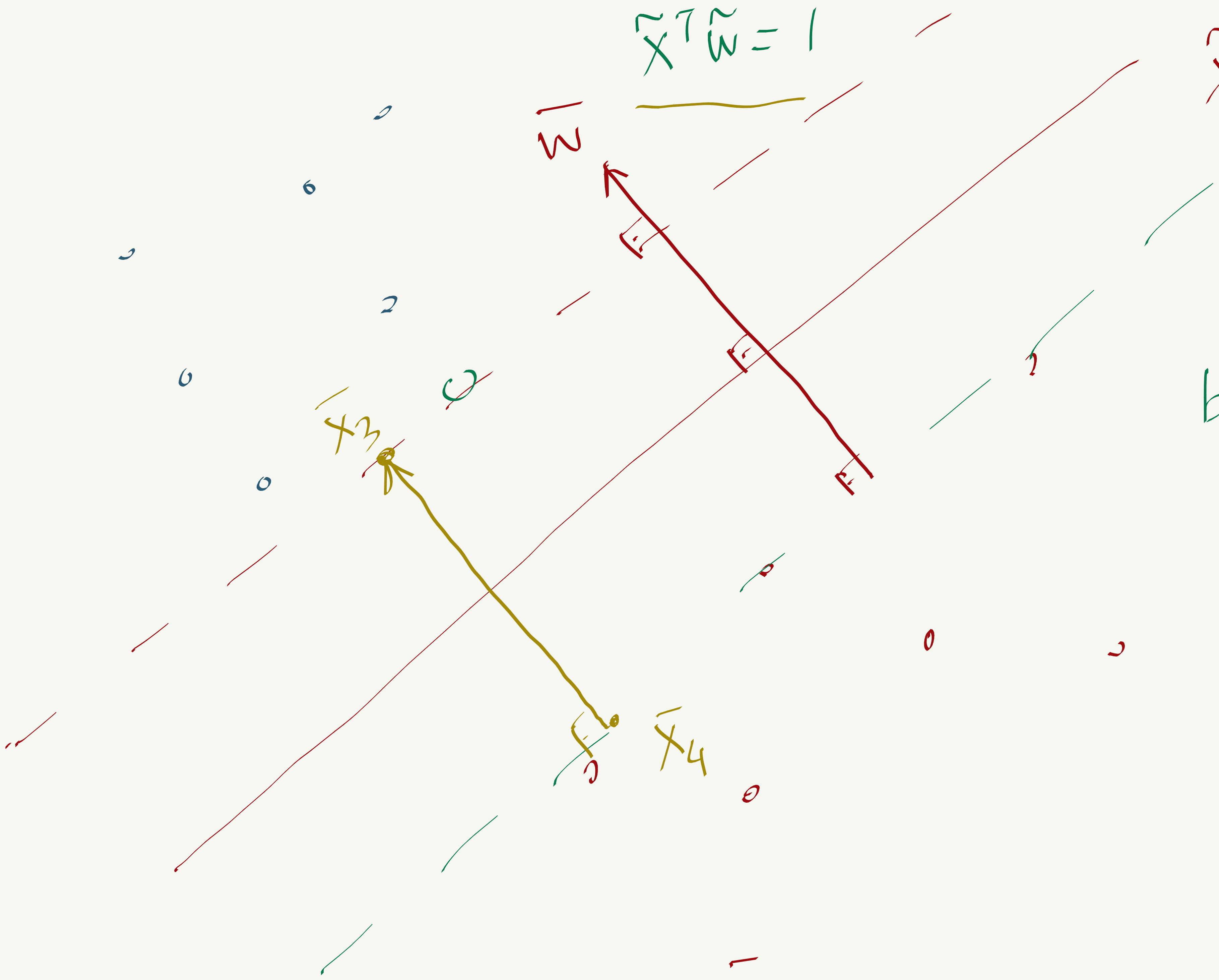
4/29/20



margin perceptron

maximum margin perceptron

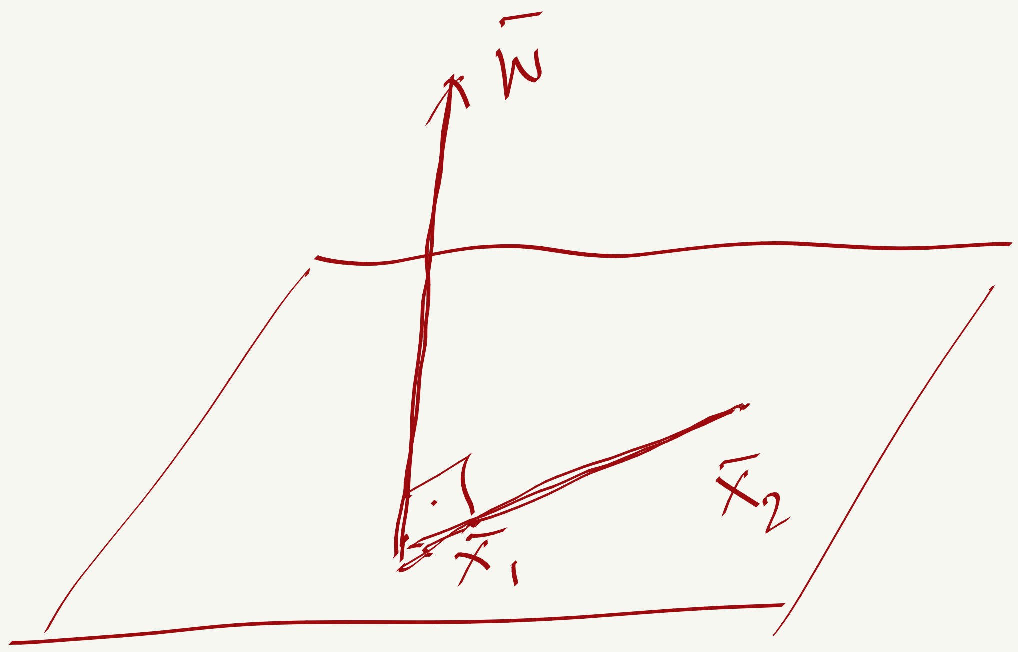
$\equiv$ Support Vector Machine (SVM) classifier



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

$\tilde{x}^T\tilde{w}=-1$

$b+\bar{x}^T\bar{w}=-1$

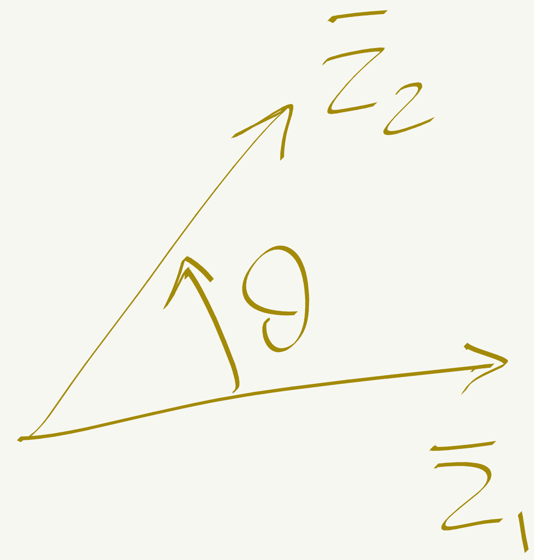


$$\left.\begin{array}{l} b+\bar{x}_1^T\bar{w}=0 \\ b+\bar{x}_2^T\bar{w}=0 \end{array}\right\} \rightarrow \underline{(\bar{x}_1-\bar{x}_2)^T\bar{w}=0}$$

$$(b+\bar{x}_3^T\bar{w})-(b+\bar{x}_4^T\bar{w})=2$$

$$\Rightarrow \underline{(\bar{x}_3-\bar{x}_4)^T\bar{w}=2}$$

$$\bar{z}_1^T \bar{z}_2 = \|z_1\| \cdot \|z_2\| \cdot \cos\theta$$



$$\Rightarrow \|\bar{x}_3 - \bar{x}_4\| \cdot \|\bar{w}\| \cdot \cancel{\cos\theta}^{\,1} = 2$$

$$\Rightarrow \|\bar{x}_3 - \bar{x}_4\| = \frac{2}{\|w\|}$$

$\therefore$ to obtain the max margin Perceptron we need to minimize $\|w\|$

Margin Perceptron

$$g(\tilde{w}) = \frac{1}{P}\sum_{p=1}^{P} \max\left(0,\ 1 - y_p\left(\tilde{x}_p^T \tilde{w}\right)\right)$$

Hard Margin SVM

$$\begin{cases} \min \boxed{\|\tilde{w}\|_2^2} \\ \text{s.t.} \\ \max\left(0,\ 1 - y_p\left(\tilde{x}_p^T \tilde{w}\right)\right) = 0, \quad \forall p \end{cases}$$

Relaxed form of the above

Soft-margin SVM

$$\min_{\tilde{w}} \left\{ \underbrace{\frac{1}{P}\sum_{p=1}^{P} \max\left(0,\ 1 - y_p \tilde{x}^T \tilde{w}\right)}_{} + \boxed{\lambda \|\tilde{w}\|_2^2} \right\}$$

Lagrange multiplier

$\boxed{\text{regularized}}$ form of margin perceptron

Approximate ReLU function

$$g(\tilde{w}) = \frac{1}{P}\sum \log\left(1 + e^{1 - y_p \tilde{x}_p^T \tilde{w}}\right) + \lambda \|\tilde{w}\|_2^2$$

ℓ_2-norm

regularized form of logistic regression

soft-max SVM

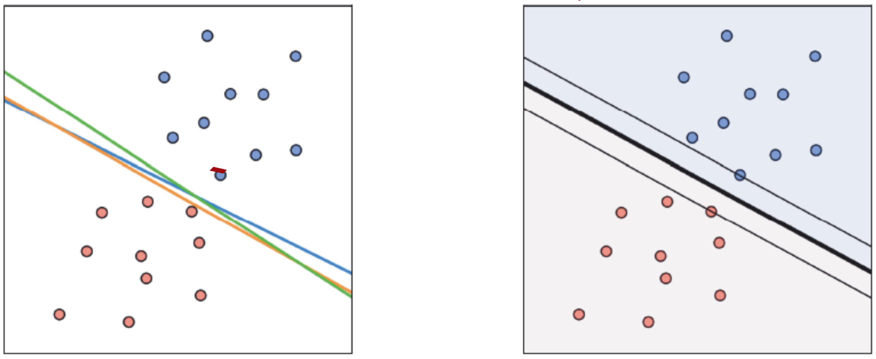
or log loss SVM

different names of

the cost function above

Same data

log regression



SVM

$\lambda=10^{-3}$

Figure 6.20 Figure associated with Example 6.8. See text for details.

Classification quality metrics

$$\text{model}(\tilde{x}, \tilde{w}^*) = \tilde{x}^T \tilde{w}$$

Prediction

given $\tilde{x}'$ $\Rightarrow$ $\text{model}(\tilde{x}'^T, \tilde{w}^*) = \tilde{x}'^T \tilde{w}^*$

$$\text{sign}(\tilde{x}'^T \tilde{w}^*) = y'$$

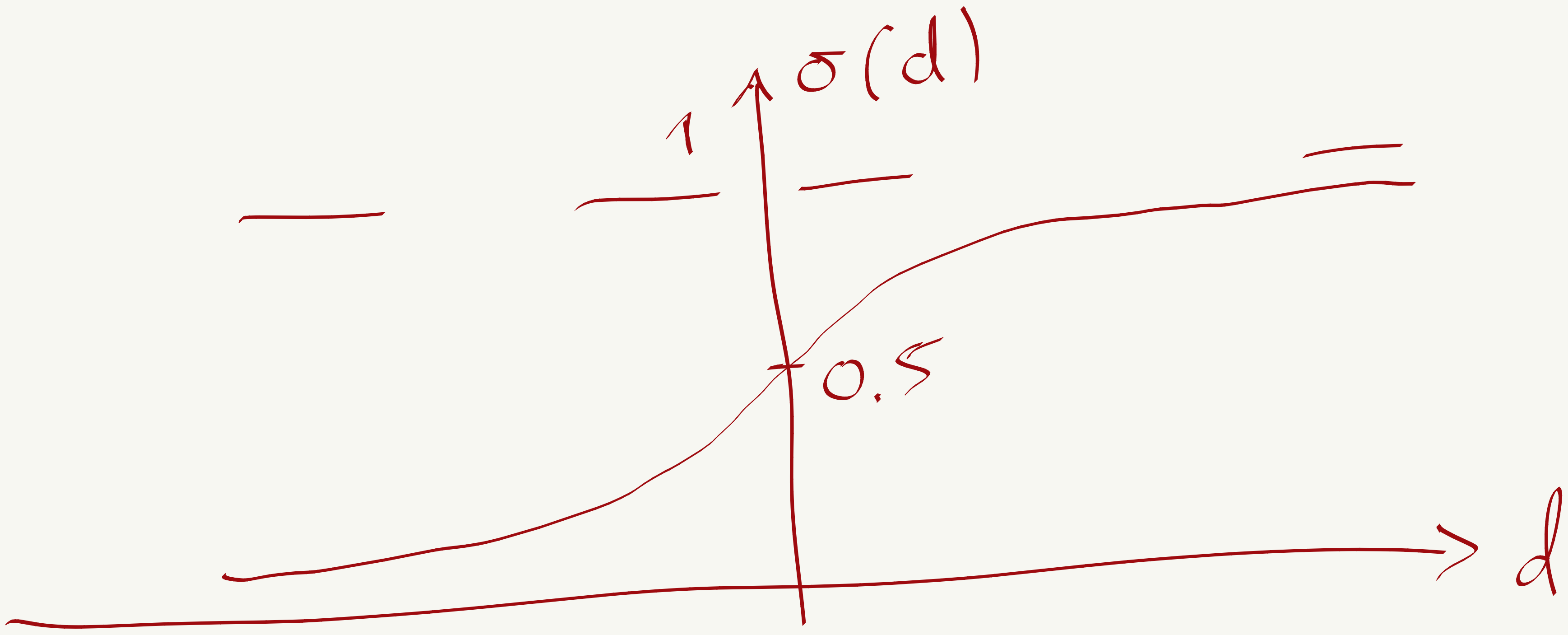
Confidence in prediction

$$(\tilde{x}' - \tilde{x}'_p)^T w = b \Rightarrow$$

$$\|\tilde{x}' - \tilde{x}'_p\| \cdot \|\bar{w}\| = b$$

$$d \cdot \|\bar{w}\| = b \Rightarrow d = \frac{b}{\|\bar{w}\|} = \frac{b + \bar{x}'^T \bar{w}}{\|\bar{w}\|}$$

Confidence: $\sigma(d)$



$\sigma(d) > 0.5 \longrightarrow$ + side of hyperplane

$\sigma(d) < 0.5 \longrightarrow$ − " " "

Quality of trained model

Accuracy

Indicator function

$$\boxed{\mathcal{I}(\hat{y}_p, y_p) = \begin{cases} 0, & \text{if } \hat{y}_p = y_p \\ 1, & \text{if } \hat{y}_p \neq y_p \end{cases}}$$

predicted label

true lable

$$\hat{y}_p = \mathrm{sign}(\mathrm{model}(\tilde{x}_p, \vec{w}^*))$$

$$\#\text{ of misclassifications} = \sum_{p=1}^{P} \mathcal{I}(\hat{y}_p, y_p)$$

$$A = 1 - \frac{1}{P} \underbrace{\sum_{p=1}^{P} \mathcal{I}(\hat{y}_p, y_p)}$$

accuracy

$$0 \leq A \leq 1$$

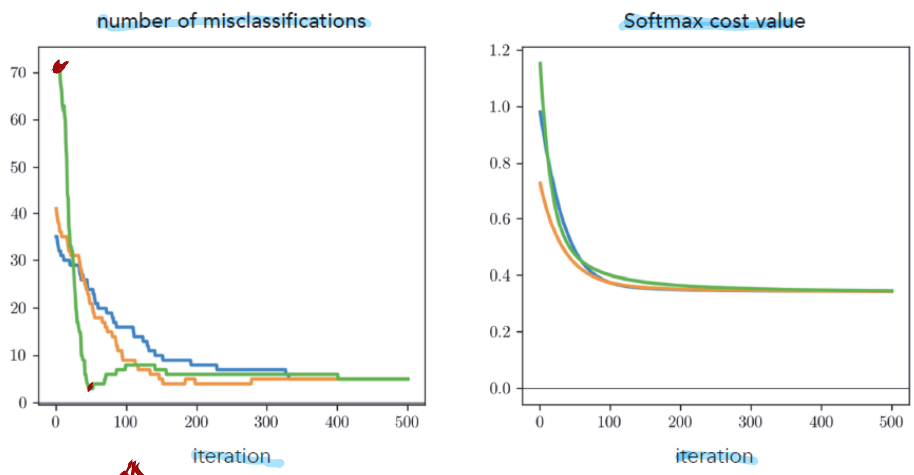


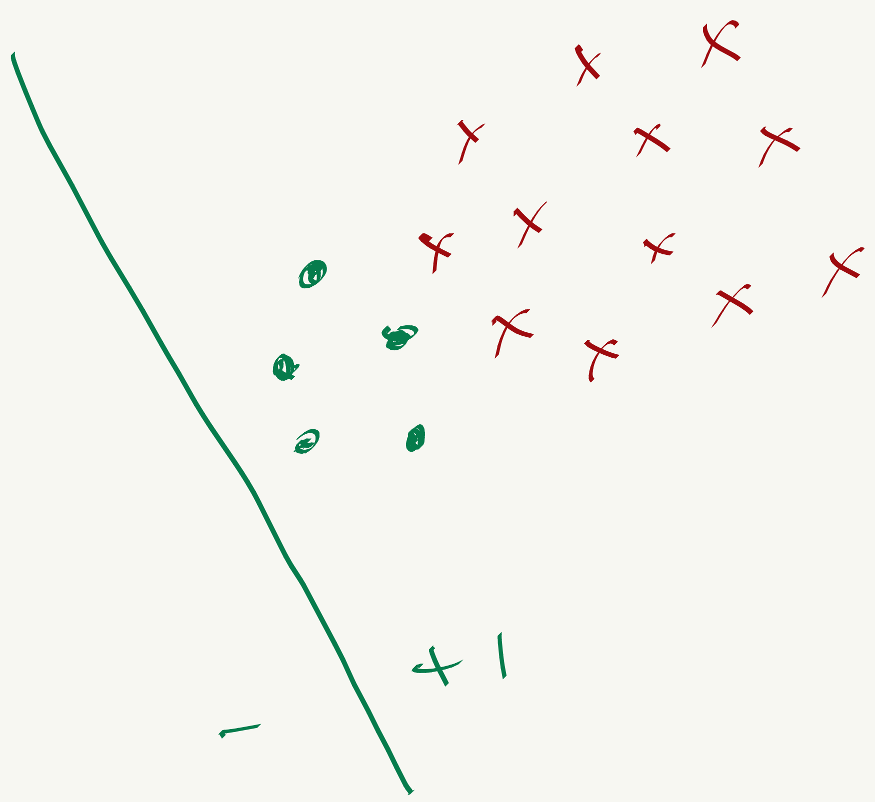
Figure 6.22 Figure associated with Example 6.9. See text for details.

metric to terminate iteration or evaluate overall performance

"Issue" w/ (global) accuracy

- if classes unbalanced $\rightarrow$ members of small class are "sacrificed"

+1 95% samples   −1 $\underline{\underline{5}}$% samples



Accuracy is 95%

<u>One solution</u> : accuracy per class

Ω_{+1} : indices of points w/ $y_p = +1$

Ω_{-1} : " " " " $y_p = -1$

$\vdots$

# of misclassification +1 class: $\sum_{p \in \Omega_{+1}} I(\hat{y}_p, y_p)$

" " -1 " : $\sum_{p \in \Omega_{-1}} I(\hat{y}_p, y_p)$

accuracy per class

$$A_+ = 1 - \frac{1}{|\Omega_{+1}|} \sum_{p \in \Omega_{+1}} I(\hat{y}_p, y_p)$$

$$A_- = 1 - \frac{1}{|\Omega_{-1}|} \sum_{p \in \Omega_-} I(\hat{y}_p, y_p)$$

$$A_{balanced} = \frac{A_+ + A_-}{2}$$

(if $|\Omega_{+1}| = |\Omega_{-1}| \Rightarrow A_{balanced} = A$)

$$0 \leq A_{balanced} \leq 1$$

Example: $|\Omega_{+1}| = 95 \Rightarrow A_+ = 1$
$|\Omega_{-1}| = 5 \Rightarrow A_- = 0$ $\Big\}$ $A_{balanced} = 0.5$