$S = (\mathscr{D}, \dot{\geq}, Q, \dot{\geq}^*, T, V, F)$

$\dot{\geq}$ - preference on document representations

$\dot{\geq}^*$ - preference on query representation

$(\mathscr{D}, \dot{\geq})$ – defines the document space

$(\mathscr{D}, \dot{\geq}^*)$ – defines the query space

ALL IN THE CONTEXT OF ONE USER NEED

Learning in Vector Model using Relevance Feedback

Optimal Query Formulation

$\underline{q} = (\underline{q}_1, \underline{q}_2 \ldots \underline{q}_n)$

$\underline{d}_\alpha = (W_{\alpha 1}, W_{\alpha 2}, \ldots W_{\alpha n})$

choose $\underline{q}$ s.t.

$$\underline{d}_\alpha \underline{q}^T = \begin{cases} \succ 0, if\ \underline{d}_\alpha \in REL \\ \prec 0, if\ \underline{d}_\alpha \in NREL \end{cases}$$

for $\alpha = 1,2,\ldots,p$

In the following we apply the "conversion" operation for $\underline{d}_\alpha$ to get $\hat{\underline{d}}_\alpha$

$$\text{Let } \hat{\underline{d}}_\alpha = \begin{cases} \underline{d}_\alpha, if\ \underline{d}_\alpha \in REL \\ -\underline{d}_\alpha, if\ \underline{d}_\alpha \in NREL \end{cases}$$

$$\hat{W} = \begin{pmatrix} \hat{\underline{d}}_1 \\ \hat{\underline{d}}_2 \\ \bullet \\ \bullet \\ \bullet \\ \hat{\underline{d}}_p \end{pmatrix}$$

$$\underline{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\hat{W} \underline{q}^T \succ \underline{0}$$

- Find the best possible $\underline{q}$
- Relevance feedback

PROBLEM STATEMENT

Find $\underline{q}$ such that
$J_p(\underline{q})$ is minimized
Subject to

$$\underline{\hat{d}}_{\alpha}\underline{q}^T \succ 0$$

for all α

This is a system of linear in equalities

Geometrical PR

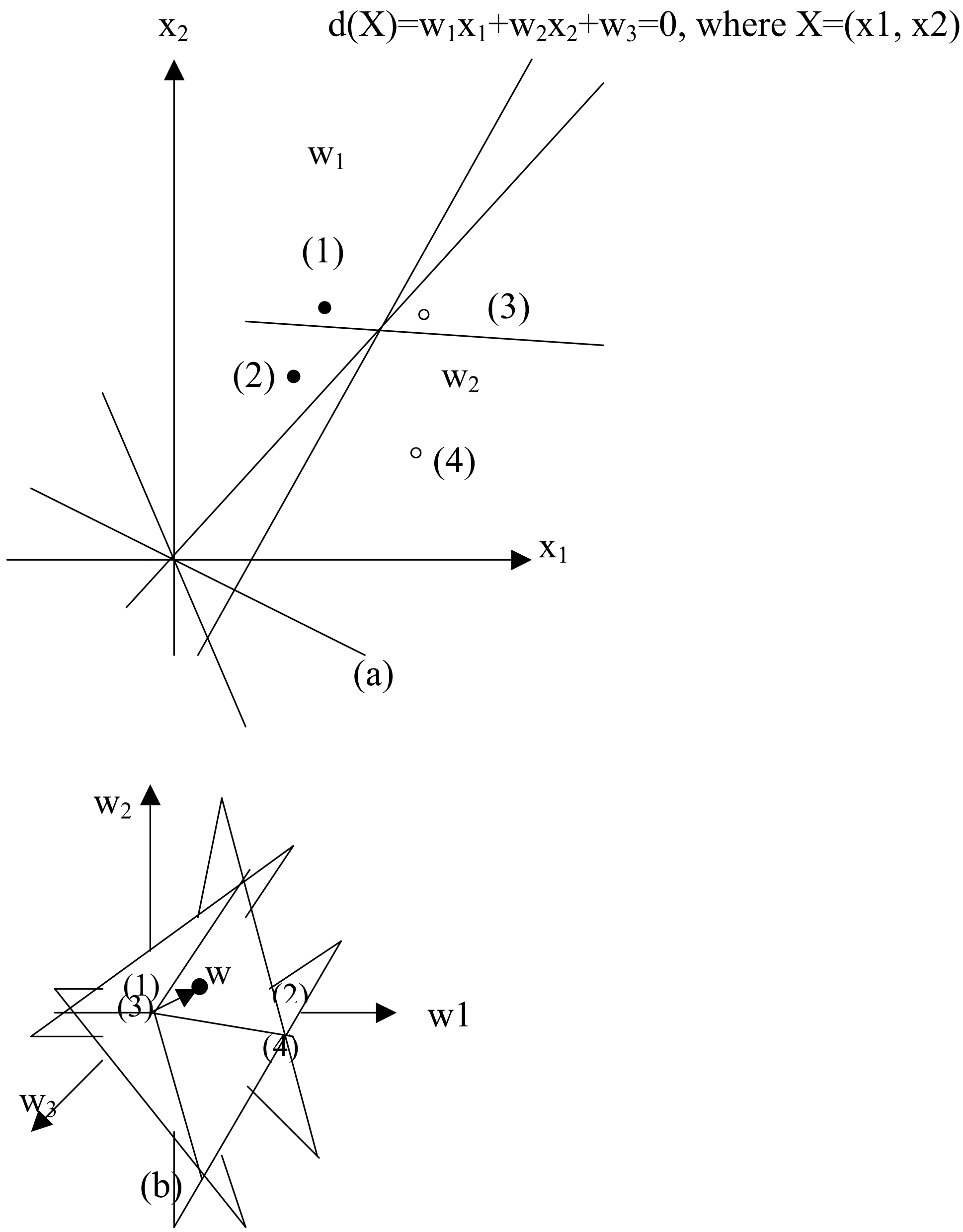
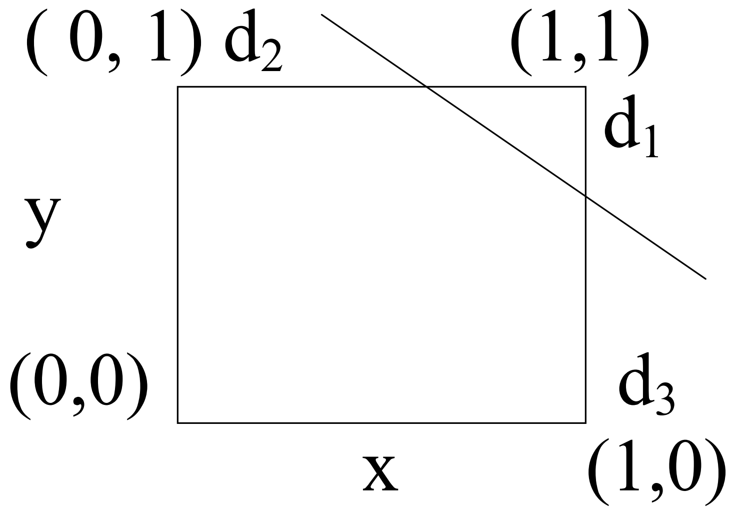


Figure 2.5. Geometrical illustration of the pattern space and the weight space.

(a) Pattern space. (b) Weight space



(1,1)	REL ← $\underline{d}_1$	
(0,1)	NREL ← $\underline{d}_2$	
(1,0)	NREL ← $\underline{d}_3$	
(0,0)	NREL	

$\hat{d}_1 = (1,1,1)$ homogeneous

$\hat{d}2 = (0,-1,-1)$ coordinate

$\hat{d}3 = (-1,0,-1)$ representation

<u>ITERATIVE (DESCENT)</u>

PROCEDURES

- Instead of a "batch" approach, here we consider an iterative approach.

$$\underline{q}^{(k+1)} = \underline{q}^{(k)} - \rho_k \frac{\partial J(\underline{q})}{\partial \underline{q}}\Big|_{\underline{q} = \underline{q}^{(k)}}$$

ρ_k – grain factor of step (iteration) k

Many forms of $J(\underline{q})$ are possible.

d_α is <u>misclassified</u> if $\hat{d}_\alpha \underline{q}^T \leq 0$

Let $y(\underline{q}) = \{\hat{\underline{d}}_\alpha \mid \hat{\underline{d}}_\alpha \underline{q}^T \leq 0\}$

Perception criterion Function

$$J_p(\underline{q}) = -\sum_{\hat{d}_\alpha \in y(\underline{q})} \hat{\underline{d}}_\alpha \underline{q}^T$$

Taking a derivative of $J_p(\underline{q})$ with respect to $\underline{q}$ yields: $\frac{\partial J_p(\underline{q})}{\partial \underline{q}} = -\sum_{\hat{d}_\alpha \in y(\underline{q})} \hat{\underline{d}}_\alpha \underline{q}^T$

We, therefore, adjust $\underline{q}(k)$ by $- \sum_{\hat{\underline{d}}_\alpha \in y(\underline{q}^k)} \hat{\underline{d}}_\alpha$

$\underline{q}^{(k+1)} = \underline{q}^{(k)} + \rho_k \sum_{\hat{\underline{d}}_\alpha \in y(\underline{q}^k)} \hat{\underline{d}}_\alpha$

$\underline{q}^k$ is modified w.r.t. a set of training samples (documents)

-- learning is by epoch

$\underline{q}^{(k+1)} = \underline{q}^{(k)} + \rho_k \underline{S}$

where $\underline{S} = \begin{cases} \underline{0}, \text{if } d_\alpha \text{ is correctly classified} \\ \hat{\underline{d}}_\alpha, \text{otherwise} \end{cases}$

-- learning is by sample

$0 < \rho_k \leq 1$

PERCEPTRON CRITERION

$$J_p(\underline{q}) = \sum_{\underline{\hat{d}} \in y(\underline{q})} \underline{\hat{d}}\,\underline{q}^T)$$

where y(q) is the set of training vectors misclassified by q

Training by Epoch

PERCEPTRON CONVERGENCE ALGORITHM (PCA)

(1) k=0; Choose initial query vector $\underline{q}^0$

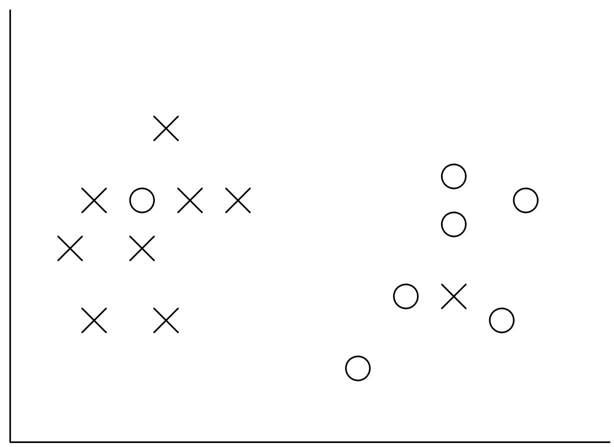
(2) Determine y ($\underline{q}^k$)

(3) If y ($\underline{q}^k$) = ϕ, terminate

(4) $\underline{q}^{k+1} = \underline{q}^k + \rho_k \sum_{\underline{\hat{d}} \in y(\underline{q}_k)} \underline{\hat{d}}$

(5) k= k+1 , go to step (2)

linearizability



Theorem: PCA will terminate if linearizability property holds for the training data. EX.

$\underline{d}_1 = (1,1,0,1,1)$ } NREL

$\underline{d}_2 = (1,0,1.0,1)$

$\underline{d}_3 = (0,1,1,0,1)$ } REL

$\underline{d}4 = (0,1,0,1,1)$ } NREL

$\hat{\underline{d}}_1 = (-1,-1,0,-1)$

$\hat{\underline{d}}2 = (1,0,1,0,1)$

$\hat{\underline{d}}3 = (0,1,1,0,1)$

$\hat{\underline{d}}4 = (0,-1,0,-1)$

$$\underline{q}^0 = (0,0,0,0,0)$$

$$y(\underline{q}^0) = (\text{all of them})$$

$$\underline{q}^1 = (0,-1,2,-2,0)$$

let $\rho_k = 1$

$$y(\underline{q}^1) = \phi$$

$$\underline{q}^{opt} = (0,-1,2,-2,0)$$