Generalization of Perception convergence Algorithm (Generalized PCA)

- use of preference
- multi-level relevance

Let documents and queries be represented as n-dimensional vectors. D is the set of all document representations. Let $<\cdot$ represent a user preference relation over D.

Defn. 1: A preference relation, $<\cdot$ is linear if there exists a query vector $\underline{q}$ such that for any $\underline{d}, \underline{d}^1 \in D$,
$\underline{d} <\cdot \underline{d}^1 \Leftrightarrow \underline{d}\underline{q}^T < \underline{d}^1\underline{q}^T$

If an algorithm is available to find such a $\underline{q}$, then we can use $\underline{q}$ to achieve <u>perfect ranking</u>.

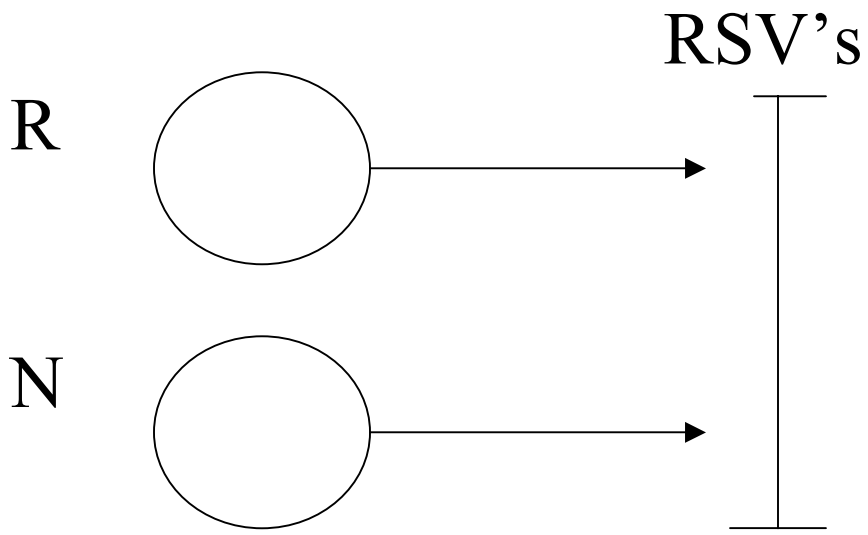
Perfect ranking is difficult to achieve and <u>not</u> <u>necessary</u>.

Defn. 2: A preference relation, $<\cdot$ is <u>weakly</u> <u>linear</u> if there <u>exists</u> a query vector $\underline{q}$ such that for any $\underline{d}, \underline{d}^1 \in D$, $\underline{d} <\cdot \underline{d}^1 \Rightarrow \underline{d}\underline{q}^T < \underline{d}^1\underline{q}^T$

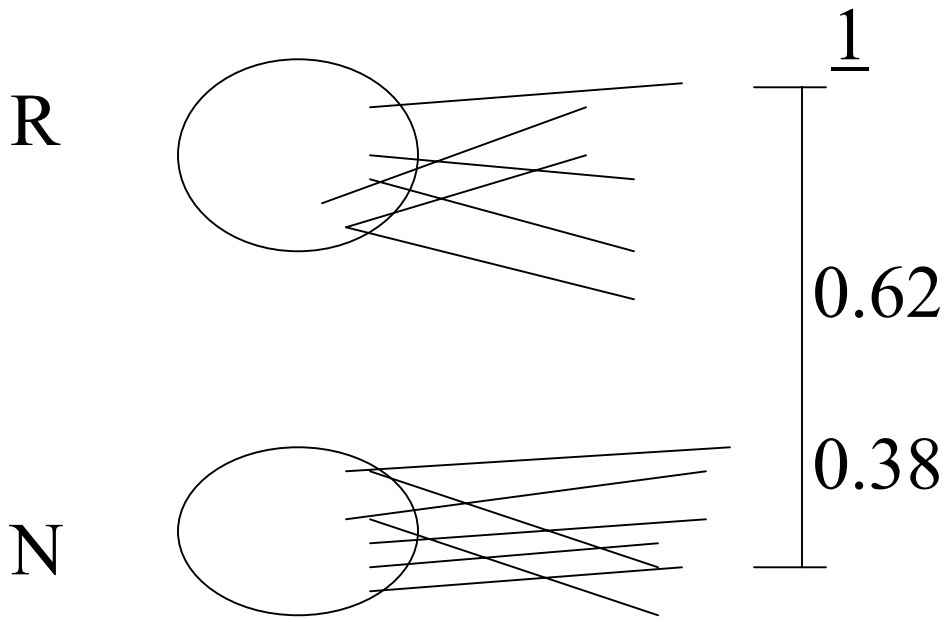
-- This condition implies $\underline{d}\underline{q}^T < \underline{d}^1\underline{q}^T \Rightarrow \text{not } (\underline{d} <\cdot \underline{d}^1)$

TWO CLASS RELEVANCE ASSUMED IN THE FIGURES BELOW

PERFECT RANKING



ACCEPTABLE RANKING



--That is, less preferable documents will not be ranked higher or equal to more preferable documents. But equally preferred documents may have different RSVs.

--If an algorithm is available to find such a $\underline{q}$, then $\underline{q}$ can be used to achieve <u>acceptable</u> <u>ranking</u>.

<u>APPROACH</u>

Suppose $\underline{d} <\cdot \underline{d}^1$ holds for $\underline{d}$, $\underline{d}^1$ in the training set. Let training set be such that user preference defined by $<\cdot$ is weakly linear. By definition,

$\underline{d} <\cdot \underline{d}^1 \Rightarrow \underline{d}\ \underline{q}^T < \underline{d}^1 \underline{q}^T$

OR

$\underline{d} <\cdot \underline{d}^1 \Rightarrow \underline{b}\ \underline{q}^T > 0$,

where $\underline{b} = \underline{d}^1 - \underline{d}$

Thus, we need to find $\underline{q}$ such that

$\underline{b}_\alpha\ \underline{q}^T > 0$

for all α, such that $\underline{b}_\alpha$ is difference between two vectors in the training set and first of the two is more preferred.

<u>Generalized Perception Criterion</u>

$J(\underline{q}) = \sum_{\underline{b} \in y(\underline{q})} -\underline{b}\underline{q}^T$,

Where

$y(\underline{q}) = \{\underline{b} = \underline{d}^1 - \underline{d} \mid \underline{d} <\cdot \underline{d}^1 \text{ and } \underline{b}\underline{q}^T \leq 0\}$

Generalized Perception Conv. Algorithm

1. choose $\underline{q}^0$, let k=0
2. let $\underline{q}^k$ be the query vector in iteration k.
   compute $y(\underline{q}^k)$
   If $y(\underline{q}^k)=\phi$, then
   Output $\underline{q}^k$ and terminate
3. Let $\underline{q}^{k+1}=\underline{q}^k+\rho_k \sum_{\underline{b}\in y(\underline{q}^k)} \underline{b}$
4. Let k=k+1, go to step 2

Example

$D = \{\underline{d}_1, \underline{d}_2, \underline{d}_3, \underline{d}_4\}$

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

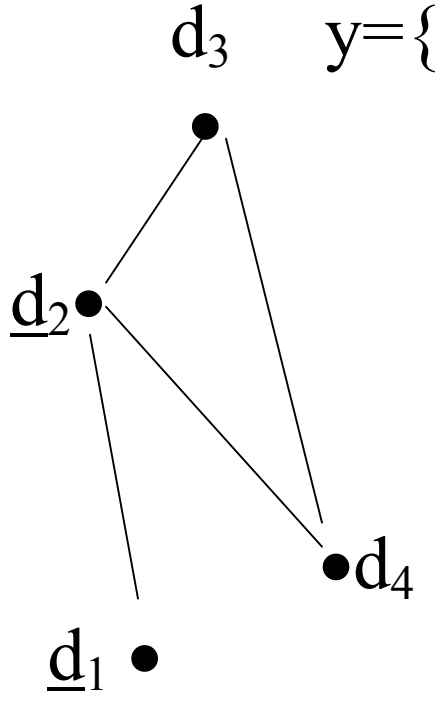
$\underline{d}_2 = (1,0,1,0)$

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

$\underline{d}_4 = (0,1,0,1)$

Let preference relation be,

$\underline{d}_1 <\cdot\ \underline{d}_2$, $\underline{d}_4 <\cdot\ \underline{d}_2$, $\underline{d}_1 <\cdot\ \underline{d}_3$, $\underline{d}_2 <\cdot\ \underline{d}_3$, $\underline{d}_4 <\cdot\ \underline{d}_3$,



$y=\{ \underline{b}_1, \underline{b}_2, \underline{b}_3, \underline{b}_4, \underline{b}_5\}$, where

$\underline{b}_1= \underline{d}_2- \underline{d}_1=(0,-1,1,-1)$

$\underline{b}_2= \underline{d}_2- \underline{d}_4=(1,-1,1,-1)$

$\underline{b}_3= \underline{d}_3- \underline{d}_1=(-1, 0,1,-1)$

$\underline{b}_4= \underline{d}_3- \underline{d}_2=(-1,1,0, 0)$

$\underline{b}_5= \underline{d}_3- \underline{d}_4=(0,0,1,-1)$

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

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

$\underline{q}^1 = \underline{q}^0 + \sum_{\underline{b}_i \in y(\underline{q}^0)} \underline{b}_i$

$= (-1,-1,4,-4)$

$y(\underline{q}^1) = \{\underline{b}_4\}$

$\underline{q}^2 = \underline{q}^1 + \underline{b}_4$

$= (-2,0,4,-4)$

$y(\underline{q}^2) = \phi$, terminate

Theorem: Generalized PCA. Will terminate if preference relation for documents in the training set is <u>weakly</u> <u>linear</u>.

<u>Example:</u>

<u>Learning by sample (Gen-PCA)</u>

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

$\underline{b}_1$ arrives → $\underline{b}_1$ is misclassified

$\underline{q}^1 = \underline{b}_1$

$\underline{b}_2$ arrives → check $\underline{b}_2\underline{q}^{1T}$

Not misclassified

$\underline{q}^2 = \underline{q}^1$

$\underline{b}_3$ arrives → check $\underline{b}_3\underline{q}^{2T}$

Not misclassified

$\underline{q}^3 = \underline{q}^2$

•

•

Rocchio's Method: (introduced in the 60's)

Cosine Similarity between a document $\underline{d}$ and query $\underline{q}$:

$$\rho(\underline{q}, \underline{d}) = \frac{\underline{d} * \underline{q}^T}{\|q\| \, \|d\|} = \frac{\underline{q}}{\|q\|} \bullet \frac{\underline{d}}{\|d\|}$$

Rocchio defined: an ideal (our definition for the same is "acceptable ranking) query q:

$$\rho(\underline{q}, \underline{d}) < \rho(\underline{q}, \underline{d}') \quad \text{for all} \quad \underline{d} \in nrel, \; \underline{d}' \in rel$$

Didn't know how to get acceptable ranking, So suggested an approximate solution. He finds the optimal query vector by maximizing the following criterion function with respect to q.

$$C = \frac{1}{n_0} \sum_{\underline{d}' \in rel} \rho(\underline{q}, \underline{d}') - \frac{1}{n_1} \sum_{\underline{d} \in nrel} \rho(\underline{q}, \underline{d})$$

where n_0. is: # of rel doc. in the set of rel.

n_1. is: # of nrel doc. in the set of nrel.

By the definition of ρ (q, d)

$$C = \frac{\underline{q}}{\|\underline{q}\|}\left[\frac{1}{n_0} \sum_{\underline{d}' \in rel} \frac{\underline{d}'}{\|\underline{d}'\|} - \frac{1}{n_1} \sum_{d \in nrel} \frac{\underline{d}}{\|\underline{d}\|}\right]$$

C is maximized if q is chosen to be:

$$\underline{q}^* = k\left[\frac{1}{n_0} \sum_{\underline{d}' \in rel} \frac{\underline{d}'}{\|\underline{d}'\|} - \frac{1}{n_1} \sum_{\underline{d} \in nrel} \frac{\underline{d}}{\|\underline{d}\|}\right]$$

because C=$\underline{q}$*•A = kA•A, where

A = $\frac{1}{n_0} \sum_{\underline{d}' \in rel} \frac{\underline{d}'}{\|\underline{d}'\|} - \frac{1}{n_1} \sum_{\underline{d} \in nrel} \frac{\underline{d}}{\|\underline{d}\|}$ and k is a positive constant. (i.e., the direction of optimal query coincides with direction of A).

The query q^* is referred to as THE optimal query by Rocchio

Problem

1. The optimal query is not necessarily a solution vector that gives acceptable ranking even if the preference relation $<_\bullet$ is weakly linear.
2. The criterion function C has a maximum only if the query vector q is ~~normalized.~~

3. This method is only defined for two-level relevance.

Ex. Same example we used for PCA.
$\{d_1, d_4\}$ is rel and $\{d_2, d_3\}$ is nrel.

$k \neq 0$

$$q' = k[\frac{1}{2}(\frac{\underline{d}_2}{\|\underline{d}_2\|} + \frac{\underline{d}_3}{\|\underline{d}_3\|}) - \frac{1}{2}(\frac{\underline{d}_1}{\|\underline{d}_1\|} + \frac{\underline{d}_4}{\|\underline{d}_4\|})]$$

$$= \frac{1}{2}(\frac{(1,1,0,1)}{\sqrt{3}} + \frac{(0,1,0,1)}{\sqrt{2}}) - \frac{1}{2}(\frac{(1,0,1,0)}{\sqrt{2}} + \frac{(0,1,1,0)}{\sqrt{2}})$$

$$= \frac{1}{2}(\frac{1}{\sqrt{3}} - \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}, \frac{-2}{\sqrt{2}}, \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{2}})$$

$$= \frac{1}{2\sqrt{6}}(\sqrt{2} - \sqrt{3}, \sqrt{2}, -2\sqrt{3}, \sqrt{2} + \sqrt{3})$$

From the Generalized PCA:

Choose $q^0 = 0$, at the first iteration,

$$y(\underline{q}^0) = \{\underline{b} = (\frac{\underline{d}'}{\|\underline{d}'\|} + \frac{\underline{d}}{\|\underline{d}\|} \mid \underline{d} \in nrel, \underline{d}' \in rel)\}$$

$$q^1 = q^0 + \sum_{\underline{b} \in y(\underline{q}^0)} \underline{b} = \sum_{\underline{d}' \in rel} \sum_{\underline{d} \in rel} (\frac{\underline{d}'}{\|\underline{d}'\|} - \frac{\underline{d}}{\|\underline{d}\|}) = n_0 n_1 [\frac{1}{n_0} \sum_{\underline{d}' \in rel} \frac{\underline{d}'}{\|\underline{d}'\|} - \frac{1}{n_1} \sum_{\underline{d} \in nrel} \frac{\underline{d}}{\|\underline{d}\|}] \approx q^*$$

⇨ Rocchio's method is the result obtained after first iteration of generalized PCA. (Here we assume: (i) training vectors are normalized (ii) the user is giving feedback using 2-level relevance)

Cycle Vs. Iteration
(of Relevance Feedback)

Training Sample. I:

Perceptron: Rocchio's:

k=0

k=1 iteration

k=2

$\hat{q}^1$

Cycle

….

Converges

$\hat{q}^1$

Training Sample II

Perceptron Rocchio's

k=0

k=1 iteration

k=2

$\hat{q}^2$

Cycle

….

Converges

$\hat{q}^2$