Rocchio's Method

More formally, Rocchio's method computes a classifier $\boldsymbol{c}_i = \langle c_{1i}, \cdots, c_{|T|i} \rangle$ for the category c_i (T is the vocabulary, that is the set of distinct terms in the training set) by means of the formula:

$$c_{ki} = \beta \cdot \sum \frac{\omega_{kj}}{|POS_i|} - \gamma \cdot \sum \frac{\omega_{kj}}{|NEG_i|}$$

where ω_{kj} is the TF-IDF weight of the term t_k in document d_j, POS_i and NEG_i are the set of positive and negative examples in the training set for the specific class c_j, β and γ are control parameters that allow to set the relative importance of all positive and negative examples. To assign a class $\tilde{c}$ to a document d_j, the similarity between each prototype vector $\boldsymbol{c}_i$ and the document vector $\boldsymbol{d}_j$ is computed and $\tilde{c}$ will be the c_i with the highest value of similarity. *The Rocchio-based classification approach does not have any theoretic underpinning and there are guarantees on performance or convergence.*