Building a model: intuition

We would like for the vectors $\mathbf{w}_i$, $\mathbf{w}_j$, and $\mathbf{w}_k$ to be able to capture the information present in the probability ratio. If we set

$$\mathbf{w}_i^\mathsf{T} \mathbf{w}_k = \log P(w_k \mid w_i)$$

then we can easily see that

$$(\mathbf{w}_i - \mathbf{w}_j)^\mathsf{T} \mathbf{w}_k = \log P(w_k \mid w_i) - \log P(w_k \mid w_j) = \log \frac{P(w_k \mid w_i)}{P(w_k \mid w_j)},$$

ensuring that the information present in the co-occurrence probability ratio is expressed in the vector space.