

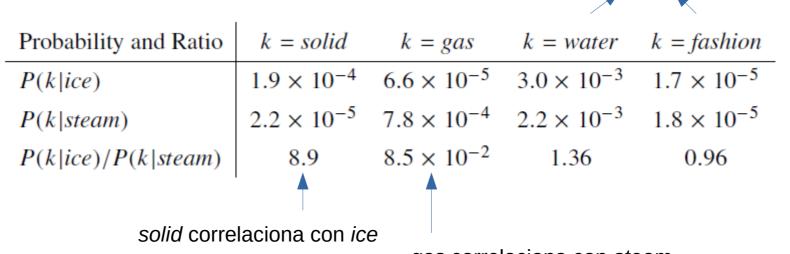
IIC 3800 Tópicos en CC NLP

https://github.com/marcelomendoza/IIC3800

- GLOVE -

- GloVe (Global Vectors):

una baja correlación da un ratio ~ 1



gas correlaciona con steam

- Fs que dependen de - :
$$F(w_i - w_j, \tilde{w}_k) = \frac{P_{ik}}{P_{jk}}.$$

- Lo expresamos vectorialmente:
$$F\left((w_i-w_j)^T\tilde{w}_k\right) = \frac{P_{ik}}{P_{jk}},$$

- Lo expresamos según F:
$$F\left((w_i-w_j)^T\tilde{w}_k\right) = \frac{F(w_i^T\tilde{w}_k)}{F(w_j^T\tilde{w}_k)}, \qquad \qquad \text{F} = \exp$$

- Consideramos que:
$$w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - \log(X_i)$$

-
$$i$$
 podría intercambiarse por k : $w_k^T \tilde{w}_i = \log(P_{ki}) = \log(X_{ki}) - \log(X_k)$

$$X_{ik} = X_{ki}$$

$$w_i^T \tilde{w}_k = w_k^T \tilde{w}_i$$

$$w_i^T w_k = w_k^T w_i$$
bias \tilde{b}_k

datos

- GloVe (Global Vectors):

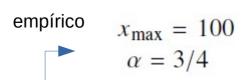
$$\frac{w_i^T \tilde{w}_k + b_i + \tilde{b}_k}{= \log(X_{ik})}$$

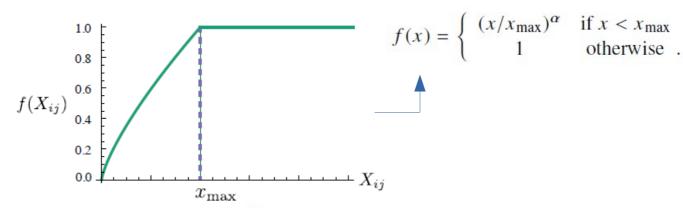
Modelo: Factorización de la matriz de co-ocurrencia

- Tarea: mínimos cuadrados.

$$\left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij}\right)^2$$

- X es sparse. Debemos compensar ese fenómeno:





- La función objetivo es:
$$J = \sum_{i,j=1}^V f\left(X_{ij}\right) \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij}\right)^2$$