Su Zhang

sz3214

Intro to Natural Language Processing - Homework 1.

① For $V=\{v_k\}_k$, consider the unigram model $p=[p_k]_{k=0}^{|V|-1}$.
where n_k = number of observations of v_k, and $p_k=\frac{n_k}{\Sigma_k n_k}$.

prove p is optimal - it maximizes the probability of the set of observations.

Step 1: The probability of a set of observed words $w_t \in V$ is defined as

$\prod_{k=0}^{|V|-1} p_k^{n_k}$ where $n_k=\sum_t w_t==v_k$. p_k is the probability assigned to word w_t.

Step 2: Take the Log of the probability.

$$\log\left(\prod_{k=0}^{|V|-1} p_k^{n_k}\right)=\sum_{k=0}^{|V|-1} n_k\cdot\log p_k$$

Step 3: the probabilities should sum to 1, which is constraint on this optimization problem.

$$\sum_{k=0}^{|V|-1} p_k=1$$

Step 4: Approach the problem through Lagrange Multipliers.

$$L(p,\lambda)=\sum_{k=0}^{|V|-1} n_k\log p_k+\lambda\left(1-\sum_{k=0}^{|V|-1} p_k\right)$$

Take the partial derivatives of the Lagrangian with respect to each p_k.

$$\frac{\partial L}{\partial p_k}=\frac{n_k}{p_k}-\lambda$$

p_k is maximized when $\frac{\partial L}{\partial p_k}=0 \Rightarrow p_k=\frac{n_k}{\lambda}$

Step 5: Solve for λ. (⇒ next page!)

$$\sum_{k=0}^{|V|-1} p_k = 1 \Rightarrow \sum_{k=0}^{|V|-1} \frac{n_k}{\lambda} = 1 \Rightarrow \lambda = \sum_{k=0}^{|V|-1} n_k$$

Step 6: Solve for p_k.

$$p_k = \frac{n_k}{\lambda} = \frac{n_k}{\sum_{k=0}^{|V|-1} n_k}.$$

therefore, the fraction of all observations that are v_k maximizes the probability of the set of observations.

Scanned with CamScanner