

## Programming Assignment-2

b03902016 資工四 周聖筌

### 1. PLSA公式推導：

Handwritten mathematical derivations for PLSA formulas:

$$\log(L(\theta)) = n(d, w) \sum_{d \in D} \sum_{w \in W} \log p(d, w).$$
$$p(z_k | w_i, d_j) = \frac{p(w_i | z_k) p(z_k | d_j)}{\sum_{k=1}^K p(w_i | z_k) p(z_k | d_j)}$$
$$p(w | z) = \frac{\sum_{d \in D} n(d, w) p(z | w, d)}{\sum_{w' \in W} \sum_{d \in D} n(d, w') p(z | w', d)}$$
$$p(z | d) = \frac{\sum_{w \in W} n(d, w) p(z | w, d)}{\sum_{w \in W} n(d, w)}.$$

### 2. 分類方法：

對於每一類，計算  $\sum_z (p(z|d) * p(w|z) \text{ for } w \text{ in } c)$  (c 是屬於某一類的字)。最後分類成數值最大的那一類。

### 3. topic number的比較：

topic number	20	30	40	50	60
public score	0.55361	0.55693	0.58229	0.57934	0.54598

越多topic number，有呈現越來越好的趨勢。但到一定數量後，再增加會導致結果變差。

### 4. 使用字典的比較：

	使用字典	不使用字典
public score	0.59153	0.58229

使用字典會上升一些。