Finding the best partition

$$L(\pi) = \underbrace{\sum_{w} n_w \log n_w}_{\substack{\text{(nearly) unigram entropy} \\ \text{(fixed w.r.t. } \pi)}} + \underbrace{\sum_{c_i, c_j} n_{c_i, c_j} \log \frac{n_{c_i, c_j}}{n_{c_i} \cdot n_{c_j}}}_{\substack{\text{(nearly) mutual information} \\ \text{(varies with } \pi)}}$$

What does maximizing this mean?

Recall

$$MI(c, c') = \sum_{c_i, c_j} p(c_i, c_j) \log \frac{p(c_i, c_j)}{p(c_i) p(c_j)}$$

which can be shown to be

$$MI(c, c') = \underbrace{\frac{1}{N}}_{\text{(constant)}} \sum_{c_i, c_j} n_{c_i, c_j} \log \frac{n_{c_i, c_j}}{n_{c_i} \cdot n_{c_j}} + \underbrace{\log N}_{\text{(constant)}}$$

and thus **maximizing MI of adjacent classes** selects the best π.