Project 2
Billy DeLucia
wld217

Readme Description:

I was able to implement the MLE and EM algorithms with good efficiency. With fairly short runtimes I was able to do multiple runs on the large corpus prior to submission. However I am not seeing the performance improvement that would be expected by using the large corpus. Although my initial performance values are not bad ~76%-81% accuracy they do not improve at all when using the larger corpus. This could either be due to an error in the in my implementation or due to the fact that the verbiage used in the larger corpus is significantly different than that of the test and training corpuses. Either way I am pleased with my results, they might not be the most accurate but they demonstrate a clear understanding of the concepts and I was able to implement them in an efficient manner.

Below you will find my proofs and figures. Since I was able to make full runs through the larger corpus I included all the values from that.

Prove that $\hat{\alpha}_t(j) = \prod_{s=1}^{t} c_s \alpha_t(j)$

$$\tilde{\alpha}_t(j) = \alpha_t(j) \cdot c_{t-1}$$

$$\hat{\alpha}_2(j) = c_2 \tilde{\alpha}_2(j) = c_2 \sum_i \hat{\alpha}_1(i)\, a_{ij}\, b_{j o_t}$$

$$= c_2 \sum c_1 \tilde{\alpha}_1(i)\, a_{ij}\, b_{j o_2} = c_1 c_2 \sum \alpha_1(i)\, a_{ij}\, b_{j o_2}$$

$$\alpha_t(j) = \sum_i \alpha_{t-1}(i)\, a_{ij}\, b_{j o_t}$$

Base Case

$$\hat{\alpha}_2(j) = c_1 c_2 \alpha_2(j) = \prod_{i=1}^{2} c_i \alpha_2(j)$$

induction

$$\hat{\alpha}_k(j) = c_k \sum_i \hat{\alpha}_{k-1}(i)\, a_{ij}\, b_{j o_k}$$

$$\therefore \hat{\alpha}_t(j) = \prod_{s=1}^{t} c_s \alpha_t(j)$$

$\sum_j \alpha_T(j) = \frac{1}{\prod_{s=1}^{T} c_s}$

$$\sum_j \alpha_T(j) = \frac{1}{\prod_{s=1}^{T} c_s}$$

$$c_t = \frac{1}{\sum_i \tilde{\alpha}_t(i)}$$

$$\tilde{\alpha}_t(j) = \sum_i \hat{\alpha}_{t-1}(i)\, A_{ij}\, B_{j\,o_t}$$

$$\hat{\alpha}_t(j) = \prod_{s=1}^{t} c_s\, \alpha_t(j)$$

$$c_t = \frac{1}{\sum_i \sum_j \prod_{s=1}^{\cancel{t}\,t-1} c_s\, \alpha_{t-1}(j)\, A_{ij}\, B_{j\,o_t}}$$

$$c_t = \frac{1}{\prod_s^{t-1} c_s \sum_i \sum_j \alpha_{t-1}(j)\, A_{ij}\, B_{j\,o_t}}$$

$$\prod_{s=1}^{t} c_s = \frac{1}{\sum_i \sum_j \alpha_{t-1}(j)\, A_{ij}\, B_j(o_t)}$$

$$\alpha_t(i) = \sum_j \alpha_{t-1}(j)\, A_{ji}\, B_{i\,o_t}$$

$$\prod_{s=1}^{t} c_s = \frac{1}{\sum_i \alpha_t(i)} \Rightarrow \sum_i \alpha_t(i) = \frac{1}{\prod_{s=1}^{t} c_s}$$

$$\therefore \sum_i \alpha_T(i) = \frac{1}{\prod_{s=1}^{T} c_s}$$

Assume $\beta_T(j) = c_T$ for all j, then prove that $\hat{\beta}_t(j) = \prod_{s=t}^{T} c_s \beta_t(j)$.

induction assumption

assume $\beta_{t+1}(i) = \left(c_{t+1} \cdots c_T\right) \beta_{t+1}(i)$

$$\hat{\beta}_t(j) = c_{t-1} \sum A_{ji} \beta$$

$$\hat{\beta}_t(j) = c_t \sum_i A_{ji} \beta_{t+1}(i) B_{j o_t}$$

$$= c_t \left[\sum_i A_{ji} [c_{t+1} \cdots c_T] \beta_{t+1}(i) B_{j o_{t+1}} \right]$$

$$= \left(c_t \cdot c_{t+1} \cdots c_T\right) \sum A_{ij} \beta_{t+1}(i) B_{i o_{t+1}}$$

Prove that $\xi_t(i,j) = \hat{\alpha}_t(i)\hat{\beta}_{t+1}(j)a_{ij}b_{jo_{t+1}}$. In other words, you can use $\hat{\alpha}$ and $\hat{\beta}$ to find ξ directly without calculating α, β and $P(\vec{O}|\lambda)$ (!).

$$\xi_t(i,j) = \frac{\alpha_t(i) A_{ij} \, B_j o_t \, \beta_{t+1}(j)}{P(o|\lambda)}$$

$$\hat{\alpha}_t(i) = \prod_{s=1}^{t} C_s \, \alpha_t(i)$$

$$\hat{\beta}_t(j) = \prod_{s=t}^{T} C_s \, \beta_t(j)$$

$$P(o|\lambda) = \prod_{s=1}^{T} C_s$$

$$\hat{\alpha}_t(i) A_{ij} \, B_i o_t \, \hat{\beta}_{t+1}(j) = \frac{\boxed{\prod_{s=1}^{t} C_s} \, \alpha_t(i) \, A_{ij} \, B_j o_t \, \boxed{\prod_{s=t+1}^{T} C_s \, \beta_{t+1}(j)}}{\boxed{\prod_{s=1}^{T} C_s}}$$

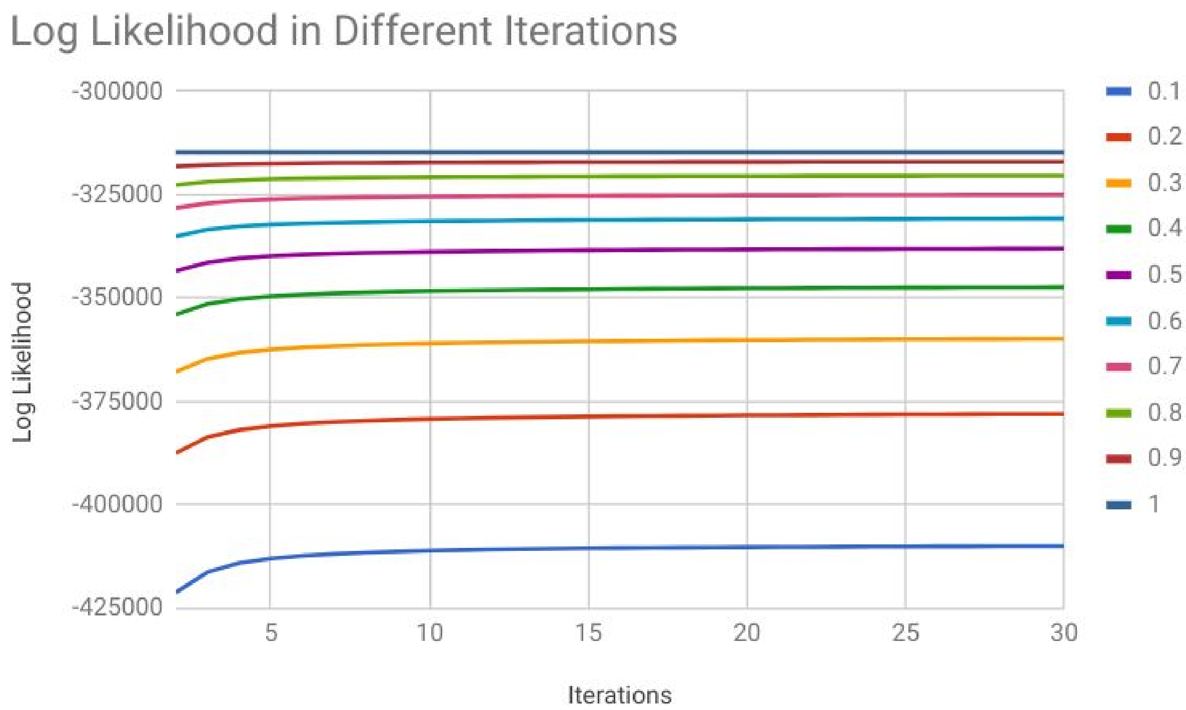
$$\therefore \xi_t(i,j) = \hat{\alpha}_t(i) \, A_{ij} \, B_j(o_t) \, \hat{\beta}_{t+1}(j)$$

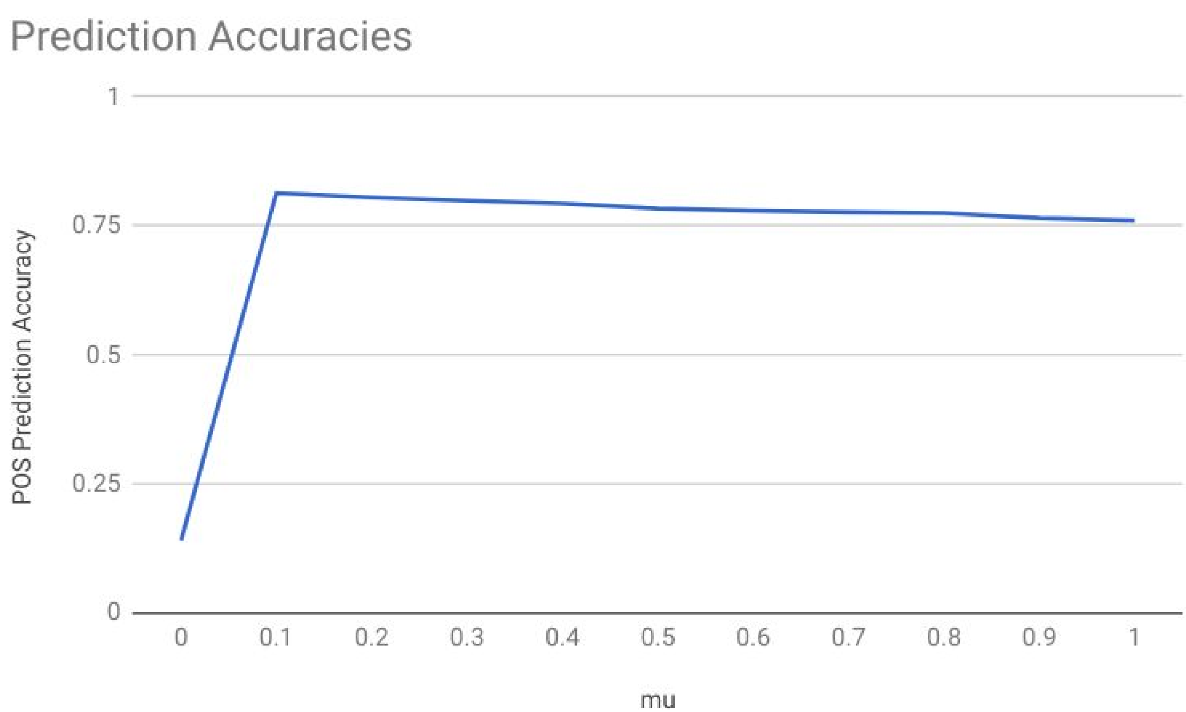
Prove that the A, B, π matrices produced by Eq. (2) define appropriate probability distributions.

Assuming that A_L, A_D, B_L, B_D, π_L, and π_D are all probabilistic distributions, the combining by weighted average of μ will result in an appropriate probability distribution

if X and Y are probabilistic distribution then $Z = \mu X + (1-\mu)Y$ will result in Z being a probabilistic distribution for all $0 \leq \mu \leq 1$

Small Corpus Data Exploration





Large Corpus Data Exploration