Algorithm Input::

- a vocabulary of size V.
- a corpus of text.
- target vector dimension D.
- context window size $2l$
- a guassian distribution for Dirichlet Process H.

Algorithm Output::

- Each word a multi-model guassian distriution.

Algorithm process::

- Initialization. Init each word's context vector c, and a mean vector μ and an isotropic variance matrix σ.

- Iteration. Move the context, and by maximizing the ojbects function,update the parameter.
  Suppose the current context center is w_i,the whole context window is $w_{i-l}, w_{i-l+1}, \ldots, w_i, \ldots, w_{i+l-1}, w_{i+l}$, the corresponding context vector is $c_{i-l}, c_{i-l+1}, \ldots, ci, \ldots, c_{i+l-1}, c_{i+l}$.
  Suppose the current context center word w_i has a representation:

  $$f(w_i) = p_1 N(\mu_1, \sigma_1) + p_2 N(\mu_2, \sigma_2) + \ldots + p_k N(\mu_k, \sigma_k) \tag{1}$$

  Denote $F(\mu_k)$: the number of times cur word belongs to a specfic model.

  $$p_i = \frac{F(\mu_i)}{\sum_{i=1}^{k} F(\mu_i)} \tag{2}$$

  mean context vector C_i

  $$C_i = \frac{\sum_{k=i-l}^{i-1} c_k + \sum_{k=i+1}^{i+l} c_k}{2 * l} \tag{3}$$

  mean target vector T_i

  $$T_i = \frac{\sum_{j=1}^{k} (p_i * \mu_i)}{k} \tag{4}$$

  Compute the Chinese RestaurantProcess hyperparameter α

  $$\alpha = \frac{T_i * C_i}{|T_i| * |C_i|} \tag{5}$$

sample from the Chinese Restaurant Process:

$$\mu \sim DP(\frac{1}{\alpha}, H) \tag{6}$$

Object function

$$L_\theta(w, c, c') = max(0, m - \log E_\theta(w, c) + \log E_\theta(w, c') \tag{7}$$

c is the positive sample,c' is the negative sample.

$$E_\theta(w, c) = \int_{-\infty}^{\infty} f(w) * f(c) \tag{8}$$

suppose $f(x) = \sum_{i=1}^{M} p_i N(\mu_i, \sigma_i)$, $g(x) = \sum_{j=1}^{N} q_j N(\mu_j, \sigma_j)$.
then

$$\int_{-\infty}^{\infty} f(x) * g(x) = \int_{-\infty}^{\infty} \sum_{i=1}^{M} \sum_{j=1}^{N} p_i * q_j N(\mu_i, \sigma_i) N(\mu_j, \sigma_j) \tag{9}$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{N} p_i * q_j \int_{-\infty}^{\infty} N(\mu_i, \sigma_i) N(\mu_j, \sigma_j) \tag{10}$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{N} p_i * q_j N(0; \mu_i - \mu_j, \sigma_i + \sigma_j) \tag{11}$$

suppose $A = \sum_{i=1}^{M} \sum_{j=1}^{N} p_i * q_j N(0; \mu_i - \mu_j, \sigma_i + \sigma_j)$,then

$$\frac{\partial \log E(f, g)}{\partial \theta} = \frac{1}{A} * \sum_{i=1}^{M} \sum_{j=1}^{N} p_i * q_j * \frac{\partial N(0; \mu_i - \mu_j, \sigma_i + \sigma_j)}{\partial \theta} \tag{12}$$

$$\frac{\log \partial E(f, g)}{\partial \mu_i} = \frac{p_i}{A} \sum_{j=1}^{N} q_j * N(0; \mu_i - \mu_j, \sigma_i + \sigma_j) * (-(\sigma_i + \sigma_j)^{-1}(\mu_i - \mu_j)) \tag{13}$$

$$\frac{\log \partial E(f, g)}{\partial \mu_j} = \frac{q_j}{A} \sum_{i=1}^{M} p_i * N(0; \mu_i - \mu_j, \sigma_i + \sigma_j) * ((\sigma_i + \sigma_j)^{-1}(\mu_i - \mu_j)) \tag{14}$$

$$\frac{\log \partial E(f, g)}{\partial \sigma_i} = \frac{1}{2} * \frac{p_i}{A} \sum_{j=1}^{N} q_j * N(0; \mu_i - \mu_j, \sigma_i + \sigma_j)((\sigma_i + \sigma_j)^{-1} * (\mu_i - \mu_j)^T * (\mu_i - \mu_j) * (\sigma_i + \sigma_j)^{-1} - (\sigma_i + \sigma_j)^{-1}) \tag{15}$$

$$\frac{\log \partial E(f, g)}{\partial \sigma_j} = \frac{1}{2} * \frac{q_j}{A} \sum_{i=1}^{M} p_i * N(0; \mu_i - \mu_j, \sigma_i + \sigma_j)((\sigma_i + \sigma_j)^{-1} * (\mu_i - \mu_j)^T * (\mu_i - \mu_j) * (\sigma_i + \sigma_j)^{-1} - (\sigma_i + \sigma_j)^{-1}) \tag{16}$$

** ******************one modal******************

suppose $f(x) = N(\mu_1, \sigma_1)$,$g(x) = N(\mu_2, \sigma_2)$ then

$$E(f,g) = \int_{-\infty}^{\infty} f(x) * g(x) \tag{17}$$

$$= N(0; \mu_1 - \mu_2, \sigma_1 + \sigma_2) \tag{18}$$

$$\log E(f,g) = -\frac{1}{2}\log\det(\sigma_1+\sigma_2) - \frac{1}{2}(\mu_1-\mu_2)^T(\sigma_1+\sigma_2)^{-1}(\mu_1-\mu_2) - \frac{d}{2}\log 2\pi \tag{19}$$

$$\frac{\partial \log E(f,g)}{\partial \mu_1} = -\frac{\partial \log E(f,g)}{\partial \mu_2} = -\Delta \tag{20}$$

$$\frac{\partial \log E(f,g)}{\partial \sigma_1} = \frac{\partial \log E(f,g)}{\partial \sigma_2} = \frac{1}{2}(\Delta\Delta^T - (\sigma_1+\sigma_2)^{-1}) \tag{21}$$

$$\Delta = (\sigma_1+\sigma_2)^{-1}(\mu_1-\mu_2) \tag{22}$$

****