$$\tilde{\Phi} = \begin{vmatrix} \Omega(\theta, \rho) & \Omega(\phi, \rho) \\ \theta & \boxed{\phi} \end{vmatrix}$$

(θ_1, ρ_1). (θ_2, ρ_2).

$$BDT^{(\theta_2,\rho_2)}\ BDT^{(\theta_1,\rho_1)}\ \phi = \begin{vmatrix} \Omega(\theta_2, \rho_2) & \Omega(\theta_1, \rho_1) & \Omega(\phi, \rho_2) \\ \Omega(\theta_2, \rho_2) & \Omega(\theta_1, \rho_1) & \Omega(\phi, \rho_1) \\ \theta_2 & \theta_1 & \boxed{\phi} \end{vmatrix}$$

$\boxed{(SW)_t = W_{\tau_k}}$?

Non-negative matrix factorization 2015-4-3

M. Berry, M. Browne, A. Langville et al. Algorithms and applications for approximate nonnegative Matrix factorization. Compu. Stati & Data analysis, 2007.

$\boxed{\text{NMF}}$ Given a non-negative matrix V, find non-negative matrix factors W and H such that. $V \approx W \cdot H$.

- Cost function
  - Euclidean distance: $\|A - B\|^2 = \sum_{i,j} (A_{ij} - B_{ij})^2$
  - Divergence: $D(A \| B) = \sum_{ij} \left(A_{ij} \log \frac{A_{ij}}{B_{ij}} - A_{ij} + B_{ij} \right)$

Rmk: Divergence is not symmetric in A & B. $D(A\|B) \rightsquigarrow$ Kullback-Leibler divergence or ~~rel~~ relative entropy when $\sum_{ij} A_{ij} = \sum_{ij} B_{ij} = 1$. in which case $D(A\|B) \rightsquigarrow \sum_{ij} \left(A_{ij} \log \frac{A_{ij}}{B_{ij}} \right)$ is a weighed average of $\log \frac{A_{ij}}{B_{ij}}$.

- Problem 1. Minimize $\|V - WH\|^2$ w.r.t. W, H s.t. $W, H \geq 0$.

  Problem 2 Minimize $D(A\|B)$ wrt W, H s.t. $W, H \geq 0$

- Multiplicative update rules

$\boxed{\text{Thm 1.}}$ The ~~Eu~~ Euclidean distance $\|V - WH\|$ is nonincreasing under the update rules.

$$H_{a\beta} \leftarrow H_{\alpha\beta} \frac{(W^T V)_{\alpha\beta}}{(W^T W H)_{\alpha\beta}}, \quad W_{ba} \leftarrow W_{ba} \frac{(V H^T)_{ba}}{(W H H^T)_{ba}}$$

The Euclidean distance is invariant under these updates iff W and H are at a stationary pt of the distance. i.e. $V = WH$ or $W = H = 0$.

$\boxed{\text{Thm 2}}$ The divergence $D(V \| W)$ is ~~nonnegative~~ nonincreasing under the update rules

$$H_{\alpha\beta} \leftarrow H_{\alpha\beta} \frac{\sum_i W_{ia} V_{i\mu} / (WH)_{i\mu}}{\sum_k W_{ka}}$$

$$W_{ba} \leftarrow W_{ba} \frac{\sum_\mu H_{a\mu} V_{i\mu} / (WH)_{i\mu}}{\sum_\nu H_{a\nu}}.$$

The divergence is invariant under these updates iff W and H are at a stationary pt of the divergence.

PROOF OF CONVERGENCE

▷ Use an auxiliary function

Defn. $G(h, h')$ is an auxiliary function for $F(h)$ if the conditions

$$\underbrace{G(h, h') \geq F(h)}_{①} \qquad \underbrace{G(h, h^*) = F(h)}_{②}$$

are satisfied

Lm1. If G is an auxiliary function, then F is nonincreasing under the update

$$h^{t+1} = \arg\min_h G(h, h^t) \quad ③$$

PROOF. $F(h^{t+1}) \underset{②}{=} G(h^{t+1}, h^{t+1}) \underset{①}{\leq} G(h^{t+1}, h^t) \underset{③}{\leq} G(h^t, h^t) = F(h^t)$

Lm2. If $K(h^t)$ is the diagonal matrix

$$K_{ab}(h^t) = \delta_{ab} (W^T W h^t)_a / h_a^t$$

then

$$G(h, h^t) = F(h^t) + (h - h^t)^T \nabla F(h^t) + \frac{1}{2} (h - h^t)^T K(h^t) (h - h^t)$$

is an auxiliary function for $F(h) = \frac{1}{2} \sum_i \left(v_i - \sum_a W_{ia} h_a \right)^2 = \frac{1}{2} (v - Wh)^T (v - Wh)$

PROOF. Obviously, we have: $G(h,h) = F(h)$. Then $\underline{G(h,h^t) \geq F(h)}$ equivalent to:

$$G(h,h^t) = F(h^t) + (h-h^t)^T \nabla F(h^t) + \frac{1}{2}(h-h^t)^T K(h^t)(h-h^t)$$

$$F(h) = F(h^t) + (h-h^t)^T \nabla F(h^t) + \frac{1}{2}(h-h^t)^T W^T W (h-h^t)$$

$$\Rightarrow \underline{G(h,h^t) - F(h) = \frac{1}{2}(h-h^t)^T (K(h^t) - W^T W)(h-h^t)} \geq 0 \quad (*).$$

which is equivalent to the prove of semi-positivity of $K(h^t) - W^T W$

The proof is incorrect !! Right now I just have no idea to correct this.

But next we should minimize $G(h,h^t)$ for fixed h^t. Namely

$\underline{h^{t+1} = \arg\min_h G(h,h^t)}$. By finding the critical pt of this one

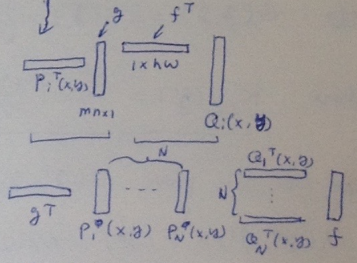
$$\nabla_h G(h,h^t) = \nabla F(h^t) + K(h^t)(h-h^t) = 0 \Rightarrow \underline{h = h^t - K(h^t)^{-1} \nabla F(h^t)}.$$

- Rethink the multiplicative update formular. 2015-4-6

The intensity at (x,y,d) pt is given by $I(x,y,d)$ $(1 \leq x \leq W, 1 \leq y \leq h)$.

Then the intensity is estimated by:

$$I(x,y) \sim \underline{\sum_{i=1}^{N} P_i g f^T Q_i} = \sum_{i=1}^{N} \underline{P_i^T(x,y)\, g\, f^T Q_i(x,y)}$$



- So we want to minimize the following function.

$$F = \sum_{x,y} \left(I(x,y) - \sum_{i=1}^{N} P_i g f^T Q_i \right)^2$$

$$= \sum_{x,y} \left(I(x,y) - g_k P_{ik}(x,y) Q_{il}(x,y) f_l \right)^2$$

⊛ Continued from last page. prove:

$$(h-h^t)^T(K(h^t)-W^TW)(h-h^t) \geq 0$$

Which is equivalent to prove:

$$v^T \operatorname{diag}(h^t)\,(K(h^t)-W^TW)\,\operatorname{diag}(h^t)\,v \geq 0.$$

Namely: $\sum_{a,b} v_a h_a^t \left(\delta_{ab} \frac{\sum (W^TW)_{a\beta} h_\beta^t}{h_a^t} - (W^TW)_{ab} \right) h_b^t v_b$

$$= \sum_{ab} \left(v_a^2 h_a^t \sum_\beta (W^TW)_{a\beta} h_\beta^t \right) - \sum_{ab} \left(v_a h_a^t (W^TW)_{ab} h_b^t v_b \right)$$

$$= \sum_{ab} v_a^2 h_a^t (W^TW)_{ab} h_b^t - \sum_{ab} v_a h_a^t (W^TW)_{ab} h_b^t v_b$$

$$= \sum_{ab} \left(\tfrac{1}{2} v_a^2 + \tfrac{1}{2} v_b^2 \right) h_a^t (W^TW)_{ab} h_b^t - \sum_{ab} v_a v_b h_a^t h_b^t (W^TW)_{ab}$$

$$= \sum_{ab} h_a^t h_b^t (W^TW)_{ab} \left(\tfrac{1}{2} v_a^2 + \tfrac{1}{2} v_b^2 - v_a v_b \right)$$

$$= \tfrac{1}{2} \sum_{ab} h_a^t h_b^t (W^TW)_{ab} (v_a - v_b)^2 \geq 0.$$

$\because$ W, $h_a^t \geq 0$

- $\min F(g) = \sum_{x,y} \left(I(x,y) - g_k P_{ik}(x,y) Q_{i\ell}(x,y) f_\ell \right)^2$ s.t. $g, f \geq 0$.

  where P_{ik}, $Q_{i\ell} \geq 0$. $\forall (x,y)$.

  <u>Defn</u>: $G(g, g^t) = F(g^t) + \sum_k (g-g^t)_k \frac{\partial F}{\partial g_k^t} + \frac{1}{2} \sum_{k,m} (g-g^t)_k K_{km}(g^t) (g-g^t)_m$

  where $K_{km}(g^t) = \delta_{km} \left(\frac{\partial^2 F}{\partial g_k^t \partial g_\beta^t} g_\beta^t \right) / g_k^t$

- Then we can prove $G(g^t, g^t) = F(g^t)$, ✓ now we prove:

  $G(g, g^t) \geq F(g)$. Namely:

  $$\sum_{km} (g-g^t)_k \left(K_{km}(g^t) - \partial_k \partial_m F(g^t) \right) (g-g^t)_m \geq 0.$$

  Which goes exactly as what did, except for here we have.

$$\frac{\partial^2 F(g^t)}{\partial g^t_k \partial g^t_m} = \frac{\partial^2}{\partial g^t_k \partial g^t_m} \sum_{x,y} \left(I(x,y) - g^t_\alpha P_{i\alpha}(x,y) Q_{i\ell}(x,y) f_\ell \right)^2$$

$$= -\frac{\partial^2}{\partial g^t_k} \sum_{x,y} 2\left(I(x,y) - g^t_\alpha P_{i\alpha}(x,y) Q_{i\ell}(x,y) f_\ell \right) P_{jm} Q_{j\ell} f_\ell$$

$$= 2 \sum_{x,y} P_{ik} Q_{i\ell} f_\ell P_{jm} Q_{js} f_s \geq 0$$

Now, $g^{t+1} = \arg\min_g G(g, g^t)$

$\min_g G(g, g^t)$ can be solved by the critical pt.

$$\nabla_g G(g, g^t) = 0 \Rightarrow \frac{\partial G(g, g^t)}{\partial g_\alpha} = \frac{\partial F}{\partial g^t_\alpha} + K_{\alpha m}(g - g^t)_m = 0$$

$$\Rightarrow g^{t+1}_\alpha = g^t_\alpha - \frac{\partial F}{\partial g^t_\alpha} K^{-1}_{\alpha\alpha}$$

$$= \frac{g^t_\alpha K_{\alpha\alpha} - \partial F / \partial g^t_\alpha}{K_{\alpha\alpha}}$$

$$K_{\alpha\alpha} = 2 \sum_{x,y} \frac{P_{i\alpha} Q_{i\ell} f_\ell P_{jm} Q_{js} f_s g^t_m}{g^t_\alpha}$$

$$\frac{\partial F}{\partial g^t_\alpha} = -2 \sum_{x,y} \left(I(x,y) - g_k P_{ik} Q_{i\ell} f_\ell \right) P_{j\alpha} Q_{j\ell} f_\ell$$

$$g^{t+1}_\alpha = g^t_\alpha + g^t_\alpha \frac{\cancel{2} \sum_{x,y} \left(I(x,y) - g^t_k P_{ik} Q_{i\ell} f_\ell \right) P_{j\alpha} Q_{j\ell} f_\ell}{\cancel{2} \sum_{x,y} P_{i\alpha} Q_{i\ell} f_\ell P_{jm} Q_{js} f_s g^t_m}$$

$$= g^t_\alpha \frac{\sum_{x,y} \left(\cancel{P_{i\alpha} Q_{i\ell} f_\ell P_{jm} Q_{js} f_s g^t_m} + \cancel{2} I(x,y) P_{j\alpha} Q_{j\ell} f_\ell \right.}{\sum_{x,y} P_{i\alpha} Q_{i\ell} f_\ell P_{jm} Q_{js} f_s g^t_m}$$

$$\left. - \cancel{2} g^t_k \cancel{P_{ik} Q_{i\ell} f_\ell P_{j\alpha} Q_{j\ell} f_\ell} \right)$$

$$= g^t_\alpha \frac{\sum_{x,y} P_{j\alpha} Q_{j\ell} f_\ell I(x,y)}{\sum_{x,y} P_{j\alpha} Q_{j\ell} f_\ell P_{jm} Q_{js} f_s g^t_m}$$

2015-4-7

- If we use the original setup:

$$\min \left\| I_d(x,y) - \sum_{i=1}^{N} P_i(x,y)\, g f^T Q_i(x,y) \right\|$$

Then we are trying to minimize the following function $F(g)$

$$F(g) = \sum_{x,y} \left(I(x,y,d) - \sum_{i=1}^{N} P_{i\alpha}(x,y)\, g_\alpha f_\ell Q_{i\ell}(x,y) \right)^2$$

$$\Rightarrow \frac{\partial F}{\partial g_k} = -2 \sum_{x,y} \left(I(x,y) - P_{i\alpha} g_\alpha f_\ell Q_{i\ell} \right)\left(P_{jk} f_s Q_{js} \right)$$ (Einstein sum convention)

$$\frac{\partial^2 F}{\partial g_k \partial g_\beta} = 2 \sum_{x,y} \left(P_{i\beta} f_\ell Q_{i\ell} \right)\left(P_{jk} f_s Q_{js} \right)$$

$$\Rightarrow K_{km}(g) = \delta_{km} \left(\frac{\partial^2 F}{\partial g_k \partial g_\beta} g_\beta \right) \Big/ g_k$$

$$= 2\delta_{km} \sum_{x,y} \frac{\left(P_{i\beta} g_\beta f_\ell Q_{i\ell} \right)\left(P_{jk} f_s Q_{js} \right)}{g_k}$$

$\Rightarrow$ Multiplicative update formula:

$$g_k^{n+1} = g_k^n - \frac{\partial F}{\partial g_k^n} \Big/ K_{kk}(g^n)$$

$$= g_k^n + \frac{\sum_{x,y} \left(I(x,y) - P_{i\alpha} g_\alpha^n f_\ell^n Q_{i\ell} \right)\left(P_{jk} f_s^n Q_{js} \right)}{\sum_{x,y} \left(P_{i\beta} g_\beta^n f_\ell^n Q_{i\ell} \right)\left(P_{jk} f_s^n Q_{js} \right)} g_k^n$$

$$= g_k^n \frac{\sum_{x,y} \left(\cancel{P_{i\beta} g_\beta^n f_\ell^n Q_{i\ell}} + I(x,y) - \cancel{P_{i\alpha} g_\alpha^n f_\ell^n Q_{i\ell}} \right)\left(P_{jk} f_s^n Q_{js} \right)}{\sum_{x,y} \left(P_{i\beta} g_\beta^n f_\ell^n Q_{i\ell} \right)\left(P_{jk} f_s^n Q_{js} \right)}$$

$$= g_k^n \frac{\sum_{x,y} I(x,y)\, P_{jk} f_s^n Q_{js}}{\sum_{x,y} \left(P_{i\beta} g_\beta^n f_\ell^n Q_{i\ell} \right)\left(P_{jk} f_s^n Q_{js} \right)}$$

- The original update formula from the paper:

$$G^{n+1} \leftarrow G^n \circ \frac{\sum_i P_i^T (W_d \circ I) Q_i^T F}{\sum_i P_i^T \left(W_d \circ \sum_j P_j G^n F^T Q_j \right) Q_i^T F}$$

$$F^{n+1} \leftarrow F^n \circ \frac{\sum_i G^{nT} P_i^T (W_d \circ I) Q_i^T}{\sum_i G^{nT} P_i^T \left(W_d \circ \sum_j P_j G^n F^T Q_j \right) Q_i^T} \qquad (10)$$