Notes on Multi-view Matrix Factor

Zhu Tianhua

09300240004

May 22, 2012

What we want to do

- Given some observed vectors $X_1, X_2, ..., X_n \in \mathbb{R}^p$ and $Y_1, Y_2, ..., Y_n \in \mathbb{R}^q$, each X_i and Y_i are features for one object.
- Reduce dimension for X's and Y's.

Ideas

- Build three matrices, P, A and Q, so that $\mathrm{E}XY^\top \approx PAQ^\top$.
- $P \in \mathbb{R}^{p\times s}, Q \in \mathbb{R}^{q\times t}, A \in \mathbb{R}^{s\times t}$
- So that $\bar{X} = P^\top X = \arg\min_{x\in\mathfrak{X}} [\mathfrak{L}_1(Px, X) + R_1(x)] \in \mathbb{R}^s$. $\mathfrak{L}_1$ is a loss function, and R_1 is a regularization parameter to encourage sparsity. $\mathfrak{X}$ is a subset of $\mathbb{R}^s$
  Similar conditions hold for $\bar{Y} = Q^\top Y$.

Relation to two-view topic model

- Consider two hidden random vectors $W \in \mathbb{R}^s$ and $Z \in \mathbb{R}^t$ that are unobserverd.
  Due to symmetry, discussions about Y is omitted here.
- W is a posibility vector which means that X is generated from a mixture W over s topics. Denote the probability of topic l as $p_l(X)$, then we have
- $p(X|W) = \sum_{l=1}^{s} W_l p_l(X)$
- $p(X|W)$ is independent of Y and Z, but W and Z are correlated.

Explaining P, A, and Q

So, we have:

$$\mathrm{E}_{X,Y} XY^\top = \sum_{l,l'} \mathrm{E} W_l Z_{l'} \int X p_l(X) \int Y^\top q_{l'}(Y) \qquad (1)$$

$P \in \mathbb{R}^{p\times s}$ satifies that $P[l,:] = \int X p_l(X)$
$Q \in \mathbb{R}^{q\times t}$, $Q[l,:] = \int Y q_{l'}(Y)$.
$A \in \mathbb{R}^{s\times t}$, $A_{l,l'} = \mathrm{E}_{W,Z} W_l Z_{l'}$

Constraints on these three matrices

$A_{i,j} \geq 0$ and $\sum_{i,j} A_{i,j} = 1$, measuring how topics are correlated.
$P, Q \geq 0. ||P[l, :]||_1 = 1, ||Q[l', :]||_1 = 1.$

Consider the following optimization:

$L_1 = \lambda_1 \mathrm{E}_X \inf_x [\mathfrak{L}_1(Px, X) + R_1(x)]$
$L_2 = \lambda_2 \mathrm{E}_Y \inf_y [\mathfrak{L}_2(Qy, Y) + R_2(y)]$
$L_0 = \mathfrak{L}_0(PAQ^\top, \mathrm{E}_{X,Y} XT^\top) + R_0(P, A, Q)$
$min_{P,A,Q}[L_1 + L_2 + L_0]$

Algorithm

$$E_{X_j,Y_k} X_j Y_k = P_j^\top A Q_k, 1 \le j \le p, 1 \le k \le q, A, P_j, Q_k \ge 0 \quad (2)$$

$$[P, A, Q] = \arg\min_{P,A,Q} \sum_{j,k} (P_j^\top A Q_k - E X_j Y_k)^2 \quad (3)$$

and $||P||_0 \le u, ||Q||_0 \le v$
Alternating least squares: Fix P, A, and solve for Q; then fix P, Q for A, and A, Q for P.
Consider diagonal A at first.

Least square

Denote $\Delta = PAQ^\top - \mathrm{E}XY^\top$, so we have
$L = \sum_{j,k}(P_j^\top AQ_k - \mathrm{E}X_j Y_k)^2 = \mathrm{trace}(\Delta\Delta^\top)$
$\frac{\partial L}{\partial \Delta} = 2\Delta$, $\frac{\partial \Delta}{\partial P} = -AQ^\top$.
Similar conclusions apply to Q.
We have two ways to optimize L. One is to use Newton's method to solve the equation, solve alternatively for the three variables:

$$L(P, A, Q) = 0 \tag{4}$$

Another method is to solve equations $\frac{\partial L}{\partial *} = 0$, where $*$ stands for P, Q and A.

Lagrange multiplier

Rewrite our target function F as:

$$F = L+R+\mu(\sum_{i,j} A_{ij}-1)+\sum_{l} \nu_l(||P[l,:]||_1 - 1)+\sum_{l'} \kappa_{l'}(||Q[l',:]||_1 - 1) \tag{5}$$

So, we have:

$$\frac{\partial F}{\partial P} = \frac{\partial(L+R)}{\partial P} + \sum_{l} \nu_l sgn(P[l,:])\mathbf{1}_l^{p\times s} = 0 \tag{6}$$

$$\frac{\partial F}{\partial Q} = \frac{\partial(L+R)}{\partial Q} + \sum_{l'} \kappa_{l'} sgn(Q[l',:])\mathbf{1}_{l'}^{q\times t} = 0 \tag{7}$$

$$\frac{\partial F}{\partial A} = \frac{\partial(L+R)}{\partial A} + \mu\mathbf{1}_{t\times s} = 0 \tag{8}$$

with all these constraints are simutaneous.

Newton's Method Take P for example.

$$\mathrm{d}L = \sum_{j,k} 2(P_j^\top AQ_k - \mathrm{E}X_jY_k)\mathrm{d}(P_j^\top AQ_k - \mathrm{E}X_jY_k)$$

$= \sum_{j,k} 2\Delta_{jk}\mathrm{d}(P_j^\top AQ_k)$
$= \sum_{j,k} 2\Delta_{jk}\mathrm{d}P_j^\top AQ_k = L$
$\Rightarrow 2\Delta_{jk}\mathrm{d}P_j^\top AQ_k = -\Delta_{jk}^2$
$\Rightarrow \mathrm{d}P_j^\top AQ_k = -1/2\Delta_{jk}$
$\Rightarrow \mathrm{d}P_j^\top = -pinv(AQ_k)\Delta_{jk}/2$
Note: here Q_k is the transpose of the k-th row in Q.

How to meet with the constraints?

Constraints.

$$A_{i,j} \geq 0, \sum_{i,j} A_{i,j} = 1 \tag{9}$$

$$P, Q \geq 0 \tag{10}$$

$$||P[l, :]||_1 = 1 \tag{11}$$

$$||Q[l', :]||_1 = 1 \tag{12}$$

The last two constraints are relatively easy to hold. We calculate 1-norm for each line of P or Q, and then divide this line by its 1-norm value, in each iteration. Such criteria will influence the accuracy (i.e., the value of L).
Initialize P and Q with $\frac{1}{pq}\mathbf{1}_{p\times s}$ and $\frac{1}{qt}\mathbf{1}_{q\times t}$, respectively, will hold the second constraint in our experiment. **Note:** $P \in [0, 1]^p$, and $Q \in [0, 1]^q$.

How to meet with the constraints? (cont'd)

$$A_{i,j} \geq 0, \sum_{i,j} A_{i,j} = 1$$

The first part, in our experiment, also holds. But the second part, for the reason that we always use A to hold the last two constraints, will not hold.

This requires us to preprocess our inputs or our model to satisfy this criterion. Here we denote λ^2 for $\sum_{i,j} A_{i,j}$, so our model should be:

$$\mathrm{E}_{X,Y} XY^\top = \lambda^2 PAQ^\top. \tag{13}$$

And when projection is applied, we use $\lambda P^\top X$ for $\bar{X}$, and $\lambda Q^\top X$ for $\bar{Y}$.