Parameterization of mixture independence models

E. Carlini

Dipartimento di Matematica
Politecnico di Torino

AMS meeting, University of Kentucky, March 2010

E.Carlini, F.Rapallo

Probability matrices, non-negative rank and parameterization of mixture models

to appear in LAA (2010)

Independence mixture models

$P \in \mathbb{R}^{IJ}$ a $I \times J$ probability matrix (i.e. non-negative with sum of the entries one) of the form

$$P = \alpha_1 c_1 r_1^t + \ldots + \alpha_k c_k r_k^t$$

where

$$c_i \in \mathbb{R}^I_{\geq 0}, \sum_j c_i(j) = 1, r_i \in \mathbb{R}^J_{\geq 0}, \sum_j r_i(j) = 1$$

and $\alpha_i \in \mathbb{R}_{\geq 0}, \sum \alpha_i = 1.$

Parameterization

$P = \sum_1^k \alpha_i c_i r_i^t$ is a probability matrix with $\mathrm{rk}(P) \leq k$.
We are describing probability matrices of rank at most k using

$$k(I+J) - k - 1$$

parameters.

Questions

Can we use fewer parameters?

What is the least possible number of parameters?

Geometry

Consider $\mathcal{M}_k \subset \mathbb{R}^{IJ}$ the variety of rank at most k matrices intersected with the simplex.

A parameterization of probability matrices with rk $\leq k$ using D parameters is an algebraic map

$$\mathbb{R}^D \supset U \longrightarrow \mathcal{M}_k$$

where U has non-empty interior.

Geometry

To obtain almost all the elements of $\mathcal{M}_k$ we need at least

$$D \geq \dim \mathcal{M}_k = k(I+J) - k^2 - 1$$

parameters.
Hence the parameterization

$$\mathbb{R}^{k(I+J)-k-1} \supset U \longrightarrow \mathcal{M}_k$$
$$(\alpha_1, \ldots, \alpha_k, c_1(1), \ldots, r_k(J)) \mapsto P = \sum \alpha_i c_i r_i^t$$

is redundant.

Non-negative rank

A matrix P such that

$$P = \sum_{1}^{k} \alpha_i c_i r_i^t$$

where $c_i \in \mathbb{R}^I_{\geq 0}, r_i \in \mathbb{R}^J_{\geq 0}$ and $\alpha_i \in \mathbb{R}_{\geq 0}$ has

non-negative rank at most k

and one writes $\mathrm{rk}_+(P) \leq k$.

Non-negative rank

The non-negative rank and the rank can be different, i.e. there are matrices P such that

$$\mathrm{rk}_+(P) \geq \mathrm{rk}(P).$$

Consider, for example

$$P = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

where $C_1 + C_2 = C_3 + C_4$ thus $\mathrm{rk}(P) = 3$, but $\mathrm{rk}_+(P) = 4$.

Non-negative rank

Hence the parameterization

$$P = \sum_{1}^{k} \alpha_i c_i r_i^t$$

is redundant and, in general, non-surjective.

Theorem *(Cohen-Rothblum)* If $\text{rk}(P) = 0, 1, 2$ then

$$\text{rk}_{+}(P) = \text{rk}(P).$$

Thus, for $k = 2$, $P = \sum_1^k \alpha_i c_i r_i^t$ is redundant but surjective.

Non-redundant parameterization for rk=2

We let $D = 2(I + J) - 5$ and we look for maps

$$\mathbb{R}^D \supset U \longrightarrow \mathcal{M}_2$$

with a dense image.

Idea:

- find $U_{j_1,j_2} \subset \mathbb{R}^D$ with non-empty interior;
- find maps $f_{j_1,j_2} : U_{j_1,j_2} \longrightarrow \mathcal{M}_2$ such that

$$\bigcup_{j_1,j_2} \mathrm{Im}(U_{j_1,j_2}) = \mathcal{M}_2.$$

Non-redundant parameterization for rk=2

We consider the map

$$f_{1,2}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \alpha) = \alpha \mathbf{a}\mathbf{b}^t + (1-\alpha)\mathbf{c}\mathbf{d}^t$$

where the vectors $\mathbf{a}, \mathbf{c}$ have sum one and the vectors $\mathbf{b}, \mathbf{d}$ have a fixed zero entry. The domain is

$$U_{1,2} = \Big\{(a_1, \ldots, a_{I-1}, b_3, \ldots, b_J, c_1, \ldots, c_{I-1}, d_3, \ldots, d_J, \alpha) \in \mathbb{R}^D :$$

$$0 \leq a_i, b_i, c_i, d_i, \alpha \leq 1 \ \text{ and } \ 0 \leq \sum a_i, \sum b_i, \sum c_i, \sum d_i \leq 1\Big\}.$$

Non-redundant parameterization for rk=2

Where $f_{1,2} : U_{1,2} \longrightarrow \mathcal{M}_2$ is defined as follows

$$f_{1,2}(a_1, \ldots, a_{I-1}, b_3, \ldots, b_J, c_1, \ldots, c_{I-1}, d_3, \ldots, d_J, \alpha) =$$

$$= \alpha \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{I-1} \\ 1 - \sum a_i \end{pmatrix} \begin{pmatrix} 1 - \sum b_i & 0 & b_3 & \ldots & b_J \end{pmatrix} +$$

$$+(1-\alpha) \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{I-1} \\ 1 - \sum c_i \end{pmatrix} \begin{pmatrix} 0 & 1 - \sum d_i & d_3 & \ldots & d_J \end{pmatrix}$$

Non-redundant parameterization for rk=2

Theorem

$$\mathrm{Im}(U_{j_1,j_2}) = \mathcal{M}_2.$$

Proof Let P be in $\mathcal{M}_2$. If $\mathrm{rk}(P) = 1$ then we get P by setting $\alpha = 0$ or $\alpha = 1$. Otherwise $\mathrm{rk}_+(P) = 2$ and we choose **a** and **c** accordingly.

Hence the maps

$$f_{j_1,j_2} : \mathbb{R}^{2(I+J)-5} \longrightarrow \mathcal{M}_2$$

give a non-redundant parameterization of the whole $\mathcal{M}_2$ and no fewer parameters can be used.

Application

Given a map

$$F : \mathcal{M}_2 \longrightarrow \mathbb{R}$$

we want maximize/minimize F, i.e. we want to solve an optimization problem on rank two probability matrices.

E.g. F can be the likelihood function and we want to solve the maximun likelihood problem.

Application

Theorem If $f_{j_1,j_2}(P') = P$ is a rank two matrix without two proportional columns, then

$$P \text{ is a maximum/minimum of } F$$

iff

$$P' \text{ is a maximum/minimum of } F \circ f_{j_1,j_2}.$$

Application

If $f_{j_1,j_2}(P') = P \in U_{j_1,j_2}$ is a rank two matrix with at least two proportional columns then

$$P' \text{ is point of the boundary of } U_{j_1,j_2}$$

but

$$f_{j_1,j_2}(P') \text{ is not a point of the boundary of } \mathcal{M}_2.$$

Hence the study of $F \circ f_{j_1,j_2}$ is not enough.

Idea to maximize F on $\mathcal{M}_2$:

- study F on rank one matrices
- maximize $F \circ f_{j_1,j_2}$ for all pairs (j_1, j_2) and make a list L of candidates extremal points
- if $P \in L$ has not two proportional columns, then $F(P)$ is a maximum
- if $P \in L$ has proportional columns it belongs to $\mathrm{Im} f_{j_1,j_2}$ for different pairs (j_1, j_2)
- we study the extremal behaviour of all the functions

$$F \circ f_{j_1,j_2}$$

  at all the point $P' \in U_{j_1,j_2}$ such that $f_{j_1,j_2}(P') = P$.
  If these behaviors agree, then $F(P)$ is a maximum, otherwise it is not.