

# Multiple Task Learning for Very Many Tasks in Very Many Dimensions

September 19, 2013

## Abstract

## 1 Introduction

**Goal:** Perform multitask learning for millions of tasks in very high dimensions.

- consider a matrix  $M \in \mathbb{R}^{TxD}$  with the rows as  $T$  task and the columns are  $D$  features or dimensions. Now the  $T$  is really big and so we need to cluster the tasks. Also, the dimension  $D$  is very large and we need to cluster the dimensions too. So it seems like we are clustering on the rows and columns of matrix  $M$ . Hence, in this case we treat the multitask learning as a co-clustering problem.
- an alternate way to achieve the above is by viewing the problem as a matrix decomposition problem where we have  $M = ABC$  and  $A \in \mathbb{R}^{TxT'}$ ,  $B \in \mathbb{R}^{T'xD'}$  and  $C \in \mathbb{R}^{D'xD}$ . Consider  $T'$  as the clustered super tasks and the  $D'$  as the clustered super-words. The matrix  $B$  that relates the task-clusters with the word-clusters is infact our relationship matrix here. In case of large number of dimensions as well as large number of features it is infeasible to work with the resulting matrix and the hope here is that a reduced dimensional or low rank representation on both the rows and columns will make the learning process much more tractable.
- we assume the super-task and the super-word structure to be latent variables and model them using LDA. Both the super-task and super-word can be learned alternately and also be used to update each other. Also, note that in the above the clustering on row-space is supervised whereas the clustering on the column space is unsupervised.

## 2 Generative Model

1. row-clustering
  - (a) each task is drawn from a distribution of super-doc
  - (b) each super-doc is drawn from a distribution of docs
2. col-clustering
  - (a) each doc is drawn from a distribution of super-words (topics)
  - (b) each super-word (topic) is drawn from a distribution of words

NOTE: the above two apparently disconnected clustering steps are linked by step 1(b) and 2(a).

## References