Statistics > Machine Learning

Learning Mixtures of Bernoulli Templates by Two-Round EM with Performance Guarantee

Adrian Barbu, Ying Nian Wu, Song Chun Zhu

(Submitted on 2 May 2013 (v1), last revised 3 May 2013 (this version, v2))

Dasgupta showed that a two-round variant of the EM algorithm can learn mixture of Gaussian distributions with near optimal precision with high probability if the Gaussian distributions are well separated and if the dimension is sufficiently high. In this paper, we generalize their theory to learning mixture of high-dimensional Bernoulli templates. Each template is a binary vector, and a template generates examples by randomly switching its binary components independently with a certain probability. In computer vision applications, a binary vector is a feature map of an image, where each binary component indicates whether a feature or structure is present or absent within a certain cell of the image domain. A Bernoulli template can be considered a statistical model for images of objects (or parts of objects) from the same category. We show that the two-round EM algorithm can learn mixture of Bernoulli templates with near optimal precision with high probability, if the Bernoulli templates are sufficiently different and if the number of features is sufficiently high. We illustrate the theoretical results by synthetic and real examples.

Comments:	11 pages, 7 figures
Subjects:	**Machine Learning (stat.ML)**
Cite as:	**arXiv:1305.0319 [stat.ML]**
	(or **arXiv:1305.0319v2 [stat.ML]** for this version)

Submission history

From: Adrian Barbu [view email]
[v1] Thu, 2 May 2013 00:31:45 GMT (1050kb,D)
[v2] Fri, 3 May 2013 09:55:22 GMT (1050kb,D)

Which authors of this paper are endorsers?