DIFFERENTIABLE UPPER BOUND FUNCTION FRAMEWORK FOR SURROGATE MIXING PROBABILITIES CONSTRUCTION IN A MODEL OF MIXTURE OF VARYING CONCENTRATIONS

V. O. Miroshnychenko[1], **T. B. Skorobohach**[2]
[1]Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[2]National Technical University of Ukraine "Kiev Polytechnic Institute", Kyiv, Ukraine
vitaliy.miroshnychenko@knu.ua, Tetianaskorobohach@gmail.com

We consider a parameter estimation problem by observation from a mixture with a finite number of components. Finite mixture models have a deep history, part of which can be read in the works [1] and [2]. Mixture models are used for sociological, medical and biological data description. There exist parametric and non-parametric models of mixtures. In study [3] a non-parametric minimax approach was applied to a finite mixture model with applications to DNA microarray data analysis.

We assume that each observed subject O belongs to one of M different mixture components ρ_m, $m = \overline{1,M}$. The sample contains n subjects $O_1, ..., O_n$. Let $\kappa_j = m$ iif $O_j \in \rho_m$. The true κ_j are unknown, but one knowns the mixing probabilities

$$p_j^m = p_{j;n}^m = P(\kappa_j = m).$$

These mixing probabilities are used to to build minimax weights $A_n = (a_j^m)_{j=\overline{1,n}}^{m=\overline{1,M}} = P_n(P_n^T P_n)^{-1}$, where $P_n = (p_j^m)_{j=\overline{1,n}}^{m=\overline{1,M}}$ and these weights are used to build various statistical models like regression models.

The D-dimensional vector of observed variables of O will be denoted by $\xi(O) = (\xi^1(O), ..., \xi^D(O))^T$, $\xi_j = \xi(O_j)$.

Let $F^{(m)}$ be distribution of $\xi(O)$ for $O \in \rho_m$,

$$F^{(m)}(A) = P\{\xi(O) \in A | O \in \rho_m\}, \text{ and } P\{\xi_j \in A\} = \sum_{m=1}^{M} p_j^m F^{(m)}(A).$$

for all Borel $A \subset R^D$.

In many problems, mixing probabilities are not observed, but they can be replaced by surrogates. In this work we are building such kind of mixing probabilities surrogate and assume their dependency on vector of parameters α. So,

$$\Big(a_j^i(\alpha)\Big)_{j=\overline{1,n}}^{i=\overline{1,M}} = A_n(\alpha) = P_n(\alpha)\Big(P_n(\alpha)^T P_n(\alpha)\Big)^{-1},$$

where $P_n(\alpha) = (p_j^i(\alpha))_{j=\overline{1,n}}^{i=\overline{1,M}}$ is a mixing probability matrix. The value $p_j^i(\alpha)$ represents the probability that the j-th element belongs to the component i given the parameters α. This vector α should be estimated.

In this talk we consider a differentiable upper bound function framework for the estimation of the parameter α and demonstrate it on classic and well known problem of mean value estimation with MVC [4] and KNN clustering [5, p. 509]. Let's remind a KNN clustering loss function, where M is a fixed number of clusters, and S_k is a set of cluster k and elements x are from sample $\Xi_n = (\xi_1, ..., \xi_n)$:

$$Loss_{knn} = \sum_{k=1}^{M} \sum_{x \in S_k} ||x - \mu_k||^2.$$

KNN clusterizer is fit by minimizing $Loss_{knn}$ over all sample splits into M clusters. Usually μ_k, $1 \le k \le n$ is mean value inside cluster S_k.

Now we use the MVC model and replace $Loss_{knn}$ by its analogue. The idea is to fit mixing probabilities parameters by minimizing this loss function:

$$Loss_a(\alpha, \mu) = \sum_{k=1}^{M} \sum_{i=1}^{n} a_j^i(\alpha) ||\xi_i - \mu_k||^2.$$

This function has no lower bound and might obtain any negative value. That is why we consider a different loss function:

$$Loss_{bound}(\alpha, \mu) = \sqrt{\sum_{k=1}^{M} \frac{1}{\lambda_k^2(\alpha)}} \sqrt{\sum_{k=1}^{M} (\sum_{i=1}^{n} p_i^k(\alpha) ||\xi_i - \mu_k||^2)^2},$$

where $\lambda_k(\alpha)$ are eigenvalues of the matrix $P_n(\alpha)^T P_n(\alpha)$.

It can be shown that

$$|Loss_a(\alpha, \mu)| \le Loss_{bound}(\alpha, \mu),$$

and minimization problem for $Loss_{bound}(\alpha, \mu)$ is much easier, especially for regression models with a large number of unknown parameters α inside mixing function $P_n(\alpha)$.

As a result, we will have following parameter estimators:

$$\hat{\mu}_k(\alpha) = \sum_{j=1}^{n} a_j^k(\alpha) \xi_j, \ \hat{\mu}(\alpha) = (\hat{\mu}_1(\alpha), ..., \hat{\mu}_M(\alpha))$$

and surrogate mixing probabilities estimator for the parameter α is defined as

$$\hat{\alpha} = \arg\min_{\alpha} Loss_{bound}(\alpha, \hat{\mu}(\alpha))$$

[1] Hall P., Titterington D., The Use of Uncategorized Data to Improve the Performance of a Nonparametric Estimator of a Mixture Density, *Journal of the Royal Statistical Society. Series B (Methodological)* **47** (1985), No. 1, 55–163.

[2] Mclachlan G., Peel D., Finite mixture models, Wiley Series in Probability and statistics, John Wiley, New York, 2000, 419+xxii pp.

[3] Maiboroda R., Sugakova O., Statistics of mixtures with varying concentrations with application to DNA microarray data analysis, *Journal of nonparametric statistics* **24** (2012), No. 1, 201–205.

[4] Maiboroda R., Kubaichuk O., Improved Estimators for Moments Constructed form Observations of a Mixture, *Theory of Probability and Mathematical Statistics* **70** (2005), 83–92.

[5] Hastie T., Tibshirani R., Friedman J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 12nd ed., Springer, 2017, 509 pp.