We thank the reviewers for their comments on the manuscript. We address a point about the separability assumption and the polynomial dependency below before giving individual responses subsequently.

Separability assumption: Note that the separability assumption is used only for the theoretical analysis; the algorithms will run and produce results even without it. To make first theoretical forays in similar setting, it is common to use idealized assumptions that give further insights. Note that, the separability assumption of the supports is not very unnatural and similar assumptions have been made for non-negative integer matrix factorization [3,12,21], clustering (Huleihel et. al. 2019) etc. Indeed, when the number of features is large and fine-grained, and ℓ is small, this condition is often satisfied. Intuitively, if the non-zero indices of the unknown sparse vectors are chosen uniformly at random for each unknown vector, then with high probability the separability assumption is true. As mentioned in L123-L141, the separability assumption is used for both the support recovery as well as the ϵ-recovery objective. Although the separability assumption can perhaps be relaxed for the support recovery, it seems unlikely that it can be relaxed for the ϵ-recovery objective.
Polynomial dependence on k and ℓ: Even to recover the support of a single k-sparse vector in the 1-bit CS setting, $\tilde{\Omega}(k^2)$ queries are required [1]. In comparison we use $\tilde{O}(k^3)$ queries for the support recovery of multiple vectors. Here a cubic dependence on k is not very unreasonable. Note that, the cubic dependence stems from the usage of the pairwise union free families that have about k^3 rows and are optimal. The dependence on ℓ is large, while the lower bound must grow at least linearly with ℓ. We agree that finding the correct lower bounds for general ℓ is an interesting open problem. For now there exists this rather large gap between the upper and lower bounds. However, ℓ is generally considered to be a constant in similar mixture models (see [23],[34]). Also, note that an exponential dependence on ℓ is sometimes unavoidable for results on sample complexities of Mixtures (see Moitra and Valiant, 2010). In this work, this large dependence on ℓ stems from the usage of generalized union free families.

Response to Reviewer 1: **"Necessity of separability"** Please see the discussion above regarding the assumptions. Note that, the assumptions are sufficient for the recovery of the vectors and are not necessary. In the example provided it is possible to recover the support of the unknown vectors since the corresponding gram matrix XX^T has this unique square-root up to a permutation. This might not always be the case. Also note that, for your example the 'complement' (0s and 1s swapped) matrix is separable, which makes the gram matrix to factorize uniquely. All this is just for support recovery, for ϵ-recovery our analysis also uses separability assumption crucially (cf. Algorithms 5,7 in supplementary).
"Simulations" Section E of the supplementary material contains some experiments on both synthetic and real world data. Since this work is mostly theoretical, we have relegated the experimental sections to the appendix.

Response to Reviewer 2: Apart from the summary the reviewer provided, we would like to point out that there are rather intricate results, such as, ϵ recovery of distinct classifiers for general ℓ, a completely non-adaptive algorithm for ϵ recovery, and also the recovery of 2 classifiers without the separability assumptions, that are not stated in the summary. While much of those proofs are in the supplementary, we urge the reviewer to take those into consideration as well. The major challenges in this work come from the de-mixing and alignment tasks, and the techniques used in this work go beyond just the usage of 1-bit compressed sensing methods, so we do not think of this as an extension of 1-bit CS.
"Separability" Please refer to the discussion on assumptions above. In the MovieLens example, genres are too crude features. In most cases more fine-grained features are used, and ℓ is small, which make the assumption more plaussible. For support recovery algorithm (Algorithm 2), the separability assumption can be relaxed slightly. Moreover, for $\ell = 2$, such an assumption is not required even for ϵ recovery.
"Complexities are large polynomials" Please see the polynomial dependence point above.
"Nonuniform mixtures" The case of non-uniform mixing weights is an interesting question. For the case of $\ell = 2$ (two unknown vectors), the exact mixing weights need not be known and we just need to know a lower bound on the smallest weight. For larger ℓ the same technique does not extend directly and will require non-trivial methods to estimate the *counts*.

Response to Reviewer 3: Thanks for appreciating the work and the constructive feedback. **"Experiments with $\ell = 2$"** Even for $\ell = 2$ case, our results are the first ones (Thm. 4, cor. 1, and the $\ell = 2$ cases of other Theorems), and therefore, no comparative results exists. Note that, [31] does not provide results of this flavor so not directly comparable.

Response to Reviewer 4: **"Dependence on k and ℓ"** Please see the polynomial dependence point above. We will elaborate this in the introduction as suggested.
"Assumption 2" This assumption is required for the correctness of Thm 4; whereas the lower bound on the magnitudes of the non-zero entries is required to determine an upper bound on value for Inf with a polynomial bit-complexity and not at all required for the correctness. That being said, both the conditions can be simultaneously satisfied as well (an element can be $o(1)$ but greater than $1/poly(n)$). We will add a discussion to clarify this in the final version.
"Clarity" (a) $\delta\mathbb{Z}^n$ refers to the set of n-dimensional vectors with entries being integer multiple of δ.
(b) The batchsize is defined in line 259 of the manuscript. We can apply Chernoff-type bounds here since the oracle answers all queries independently.