Review for
"Optimality in High-Dimensional Tensor Discriminant Analysis"

Jiaxin Hu

This work focuses on the theoretical analysis for the tensor discriminant analysis (TDA) model and the high-dimensional TDA (HD-TDA) estimator in previous work Pan et al. (2018). Several novelty and correctness concerns arise in the analyses and results, though TDA problem itself is interesting. There exists a huge room in the manuscript for improvement.

Major concerns:

1. The HD-TDA estimator error rate and proof techniques in Theorem 3.1 and 3.3 are not new in the field. Specifically, the analysis for HD-TDA relies on the formula equivalence between TDA and linear discriminant analysis (LDA) model: vectorizing TDA model (2.1) leads to a LDA model of dimension $p = \prod_{m\in[M]} p_m$ with special covariance matrix $\Sigma = \Sigma_1 \otimes \cdots \otimes \Sigma_M$, where p_m and Σ_m are dimension and covariance matrix on the m-th mode of the tensor data, respectively, and $\otimes$ is the matrix Kronecker product. The vector-based analysis ignores the benefit from the tensor model. For example, in Lemma A.1 of the Supplement, the estimation error for the less-parameterized Σ is the same for an unconstrained p-by-p matrix. The sparsity assumption is also independent with the tensor structure. As pointed out in Remark 3, authors obtain the same result as previous LDA work (Tony Cai and Zhang, 2019) using similar techniques. Therefore, it is critical to highlight the "additional work" uniquely required by the TDA analysis compared with lower-order cases.

2. The minimax result in Theorem 3.2 is also the same as previous work (Tony Cai and Zhang, 2019), regarding the conclusion and proof techniques. TDA-specific analysis should be emphasized given the formula equivalence between TDA and LDA. In addition, the minimax result for multi-class scenario is more interesting. Establishing or explaining the difficulties for multi-class minimax bound is helpful.

3. The statements for main theorems are incorrect. *First*, the signal condition for Theorem 3.2 is incorrect. Taking $s = \sqrt{p}$ and $n = p$, the parameters satisfy the conditions of Theorem 3.1 and 3.2 simultaneously, however, the conclusions in these two theorems are contradict. *Second*, the balance condition (upper and lower bounds) for n_k is missing. In Page 16 of Supplement, authors use the condition $n_k \asymp n$ directly, which is not true without the boundedness of n_k. The proof is invalid under the extreme case in which $n_1 = 1$ as n increases. *Third*, the number of class K can not be "arbitrary". When $K = n$, the sample size $n_k = 1$, and there is no hope to fulfill the balance condition. A general upper bound for K is necessary.

4. The simulation is inadequate. More experiments that verify the minimax result with varying dimensions, sample sizes, and signal levels are needed. Also, computational complexity and sensitivity analysis for model misspecification will strengthen the simulation.

Minor:

1. The subscript (j) in the formula of $\hat{\Sigma}_j^{(k)}$ on Page 4 is not defined.
2. The parameter space and upper bound results for binary and multiclass scenario are repetitive; combining these parts will be more efficient for readers.

References

Pan, Y., Mai, Q., and Zhang, X. (2018). Covariate-adjusted tensor classification in high dimensions. *Journal of the American statistical association*.

Tony Cai, T. and Zhang, L. (2019). High dimensional linear discriminant analysis: optimality, adaptive algorithm and missing data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(4):675–705.