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

Jiaxin Hu

The main contribution of the work lies in the theoretical analysis for the tensor discriminant analysis (TDA) model proposed in Pan et al. (2018). Though the TDA problem is interesting, several concerns for the novelty and theoretical results arise, and I believe there exists a huge room in the manuscript for improvement.

Major concerns:

1. Novelty concern arises in the theoretical analyses of the high-dimensional TDA (HD-TDA) estimator proposed by Pan et al. (2018), which occupies a major space in the manuscript. Note that the TDA model (2.1) is equivalent to the linear discriminant analysis (LDA) model of dimension $p = \prod_{m\in[M]} p_m$ with a special covariance matrix $\Sigma = \Sigma_1 \otimes \cdots \otimes \Sigma_M$, where p_m's are the dimensions of the tensor data and $\otimes$ is the matrix Kronecker product. The proof of main Theorem 3.1 relies on the equivalence between TDA and LDA model, and the tensor structure as well as the special structure in covariance matrix are ignored. Specifically, the Lemma A.1 does not reflect the fact that the unknown parameters in Σ is less than an unconstrained p-by-p matrix. Moreover, the benefit of special covariance structure will be covered by the estimation error in the mean tensor μ or its vectorization δ since the sparsity assumption is only addressed on the production tensor $\boldsymbol{B}$. Therefore, no tensor-specific technique is adapted to obtain the results for TDA, and it is not surprising to see the upper bound is equivalent to the LDA results (Tony Cai and Zhang, 2019). It is critical to point out what specific "additional work" is required by the proposed TDA model compared with the study for lower-order cases.

2. Novelty concern also arises for the minimax bound result in Theorem 3.2. Since the TDA model can be considered as a special case of LDA model (as concern # 1 mentioned), the result in Theorem 3.2 is not new to the field considering the previous minimax result in Tony Cai and Zhang (2019). Note that the proof of Theorem 3.2 also follows Tony Cai and Zhang (2019). It is worthwhile to point out the conceptual and technical differences between the Section 3.1.3 and Tony Cai and Zhang (2019) or other previous works for lower-order cases. In addition, it is also helpful to explain the difficulties to establish the minimax bound for the multiclass classification.

3. There may exist some incorrectness and missing assumptions in the theoretical results. *First*, the signal conditions in Theorem 3.1 and 3.2 are contradict. The Theorem 3.2 requires $\log s \asymp \log p$ where s refers to the sparsity. Taking $s = \sqrt{p}$, the parameters also satisfy the condition in Theorem 3.1 when $n = p$, whereas, the conclusions in Theorem 3.1 and 3.2 are contradict. Correction or explanation should be provided. *Second*, there is no upper and lower

bounds for n_k in all the theorem statements while the proof directly uses the assumption that $n_k \asymp n$ without explanation; see Page 16 of the Supplement. The proof will be invalid under the extreme example where $n_1 = 1$ as n increases. The balance condition on n_k or further explanation should be added. *Third*, there should have a general upper bound of the number for the number of class K, rather than "arbitrary number of classes". At least, K should be no larger than the sample size n. Also, if $K = n$, then $n_k = 1$ for all $k \in [K]$ and the assumption $n_k \asymp n$ does not hold as n increases. Therefore, there should have a speed limit for the growing of K when sample size n and dimension p increase.

4. The simulation is also not enough. Current simulation results only verify the upper bound results. More experiments should be implement to verify the minimax bound with varying dimension, sample size, and the signal level. Also, numerical results for method robustness to model misspecification and computational complexity would be helpful.

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 of binary and multiclass classfication are repetitive. Combining the upper bound results in these two cases 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.