Perturbation Methods for Discriminant Analysis

Yung-Kyun Noh[⋆], Jihun Hamm[†], and Daniel D. Lee[⋆]
[⋆]GRASP Lab., Department of Electrical and Systems Engineering,
University of Pennsylvania, Philadelphia, PA 19104
[†]Section of Biomedical Image Analysis,
University of Pennsylvania, Philadelphia, PA 19104
{nohyung, jhham, ddlee}@seas.upenn.edu

Topic: learning algorithms, **Preference: Oral/Poster**
Presenting author: Yung-Kyun Noh

Discriminant analysis seeks to find a linear subspace that preserves the discriminability information when data are projected onto the subspace. Considering the optimal discriminability of the projected distributions in the subspace, the possible criterion for dimensionality reduction is minimizing the Bayes error:

$$J_1(W) = \int \min[\pi_1 p_1(W, \mathbf{x}), \pi_2 p_2(W, \mathbf{x})] d\mathbf{x} \tag{1}$$

where π_1 and π_2 are the priors of two classes, and $p_1(W, \mathbf{x})$ and $p_2(W, \mathbf{x})$ are the likelihoods in the subspace projected by matrix W. However, finding W that directly minimizes J_1 is analytically intractable.

We can instead solve this problem by minimizing a tractable upper bound. One example is the Bhattacharyya upper bound which is analytically integrable when two likelihoods are Gaussians. In particular, we consider the Bhattacharyya coefficient which is the log of integrated Bhattacharyya bound.

$$\begin{aligned} J_2 &= \log\left[\int p_1^{1/2}(\mathbf{x}) p_2^{1/2}(\mathbf{x}) d\mathbf{x}\right] \\ &= -\frac{1}{2} Tr[S_W^{-1} S_B] - \frac{1}{2}\log \frac{|\frac{\Sigma_1+\Sigma_2}{2}|}{|\Sigma_1|^{1/2}|\Sigma_2|^{1/2}} \end{aligned} \tag{2}$$

Here, S_W and S_B are the within class covariance and between class covariance in the subspace as defined in Fisher Discriminant Analysis (FDA) (Fukunaga, 1990), and Σ_1 and Σ_2 are the projected covariance matrices of two classes. The standard way of optimizing this criterion is gradient descent, but this may be problamatic if the landscape has many local minima. Also, one cannot apply deflation with this method in order to determine each column of W one by one.

We investigate how to optimize the Bhattacharyya criterion with the same complexity of eigenvalue problem. First, we note that this criterion is composed of two separate terms which are easily solvable by themselves. One situation arises when the two Gaussians are homoscedastic. In this case, the solution is the FDA solution using just the first term of the Bhattacharyya criterion. The other situation arises when the two means are identical; in this case the analytic solution comes form optimizing the second term.

We note that optimizing the complete Bhattacharyya criterion forms two regimes whose solutions can be significantly different. Figure 1 shows how the solution transitions from FDA behavior to the other regime as the distance between the means decreases. This can be either a smooth (2nd order) or discontinuous (1st order) transition.

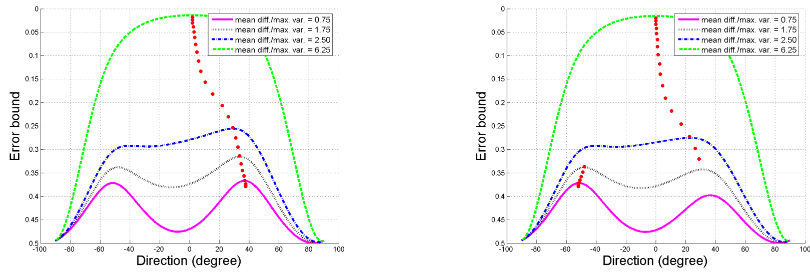

Figure 1: Bhattacharyya coefficient for 2 Gaussians in two dimensions as the distance between means vary along an angle of 0 degree. Red dots indicate global optimal projection directions.

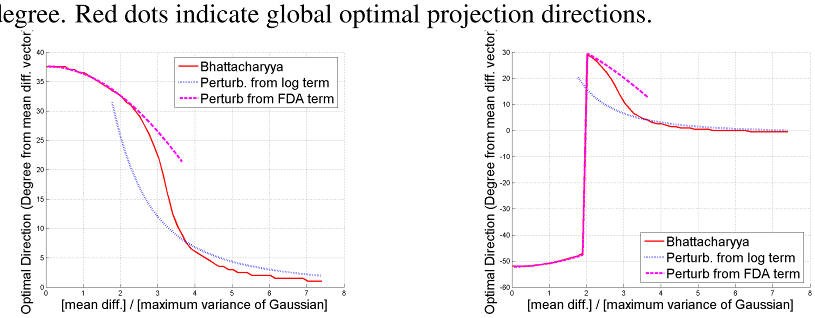

Figure 2: Analysis of optimal directions for Bhattacharrya coefficient. We show how first order perturbation influences the unperturbed direction to approximate these optimal directions.

We apply perturbation theory to find the optimal solution along this entire transition region. We show how to perturb either the analytic FDA or Fukunaga solutions to obtain an analytically tractable approximation of unsolvable problem. One example of perturbing the log term solution is shown here:

$$w_n = \phi_n + \sum_{k \neq n} \frac{4\phi_k}{E_{0n}^{(0)} - E_{0k}^{(0)}} \sum_{i=1}^{2} g_{in}^{(0)} \phi_k^T (S_B - E_{in}^{(0)} \Sigma_i) \phi_n \tag{3}$$

$$\left(g_{in}^{(0)} = \frac{\phi_n^T \Sigma_2 \phi_n}{\phi_n^T \Sigma_i \phi_n}, E_{0n}^{(0)} = \frac{\phi_n^T \Sigma_1 \phi_n}{\phi_n^T \Sigma_2 \phi_n}, E_{in}^{(0)} = \frac{\phi_n^T S_B \phi_n}{\phi_n^T \Sigma_i \phi_n} \right) \tag{4}$$

where ϕ_n is the nth eigenvector satisfies $\Sigma_1 \phi_n = \lambda_n \Sigma_2 \phi_n$.

Finally, we discuss how to extend this perturbation analysis, and compare our approach with other conventional approximations.

References

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.

K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego, California, 1990.

Stephen Gasiorowicz. *Quantum Physics*. Wiley, 2003.