Information Theory Notes

Charles Zheng and Yuval Benjamini

November 25, 2015

These are preliminary notes.

1 Classification in high-dimension, fixed SNR regime

We observe a data point y_* which belongs to one of K classes. The distribution in the ith class is $N(\mu_i, \Omega)$. We have another dataset which we use to estimate the centroids μ_i: we obtain estimates $\hat{\mu}_i \sim N(\mu_i, \Xi)$. The class centroids were originally drawn i.i.d. from a multivariate normal $N(0, \Sigma)$. Furthermore Σ, Ω, Ξ are unknown and have to be estimated as well: assume we have obtained estimates $\hat{\Sigma}, \hat{\Omega}, \hat{\Xi}$ via some method. Without loss of generality, take the case $\Sigma = I$, and take the Kth class to be the true class of y_*. Write $\hat{\mu}_* = \hat{\mu}_K$.

The classification rule is given by

$$\text{Estimated class} = \text{argmin}_i \ (y_* - B\hat{\mu}_i)^T A (y_* - B\hat{\mu}_i)$$

where A and B are matrices based on $\hat{\Sigma}, \hat{\Omega}$ and $\hat{\Xi}$. The Bayes rule is given by

$$A_{Bayes} = (I + \Omega - (I + \Xi)^{-1})^{-1}$$

$$B_{Bayes} = (I + \Xi)^{-1}.$$

The "plug-in" estimates of A and B are

$$A = (\hat{\Sigma} + \hat{\Omega} + \hat{\Sigma}(\hat{\Sigma} + \hat{\Omega})^{-1}\hat{\Sigma})^{-1}$$

$$B = \hat{\Sigma}(\hat{\Sigma} + \hat{\Xi})^{-1}.$$

Note that

$$\begin{aligned}(y_* - B\hat{\mu}_i)^T A(y^* - B\hat{\mu}_i) &= ||A^{1/2}y_* - A^{1/2}B\hat{\mu}_i||^2 \\ &= ||A^{1/2}y_*||^2 + ||A^{1/2}B\hat{\mu}_i||^2 - 2\langle A^{1/2}y_*, A^{1/2}B\hat{\mu}_i\rangle.\end{aligned}$$

If we consider the limit where $\mathrm{tr}(\Sigma^2) \to 0$, $\mathrm{tr}(\Omega^2) \to 0$ and $\mathrm{tr}(\Xi^2) \to 0$, and where the same property holds for their estimators, then the norms $||A^{1/2}B\hat{\mu}_i||^2$ converge in probability to a constant. To be specific,

$$\mathbf{E}||A^{1/2}B\hat{\mu}_i||^2 = \mathrm{tr}(AB(I+\Xi)B^T)$$

$$\mathrm{Var}||A^{1/2}B\hat{\mu}_i||^2 = 2\mathrm{tr}(AB(I+\Xi)B^T)^2.$$

(Check this later...)

Therefore, the classification rule is asymptotically equivalent to the rule

$$\text{Estimated class} = \mathrm{argmax}_i\ Z_i$$

where

$$Z_i = \langle A^{1/2}y_*, A^{1/2}B\hat{\mu}_i\rangle.$$

In the limit considered, $Z_*, Z_1, \ldots, Z_{K-1}$ are jointly asymptotically normal, with distribution

$$\begin{bmatrix} Z_* \\ Z_1 \\ \vdots \\ Z_{K-1} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathrm{tr}(AB) \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 + \mathrm{tr}(ABAB) & & & 0 \\ & \sigma^2 & & \\ & & \ddots & \\ 0 & & & \sigma^2 \end{bmatrix} \right),$$

where

$$\sigma^2 = (\mathrm{tr}AB(I+\Xi)B^T)(\mathrm{tr}A(I+\Omega)).$$