Analyse linéaire discriminante pour des modèles de variances hétérogènes

Abla KAMMOUN[1], Houssem SIFAOU[1], Mohamed-Slim ALOUINI[1]

[1]Laboratoire de la théorie de Communication
Université des sciences et technologies du Roi Abdallah Arabie Saoudite
abla.kammoun@gmail.com, houssem.sifaou@kaust.edu.sa, slim.alouini@kaust.edu.sa

Résumé – L'analyse linéaire discriminante est une technique de classification largement utilisée dans le domaine de l'apprentissage statistique. Sous l'hypothèse de distribution gaussienne, elle correspond à la méthode de classification optimale selon le principe de maximum de vraisemblance, sous condition que les classes ont une covariance commune, parfaitement connue ainsi que leurs moyennes respectives. En pratique, ces quantités statistiques ne sont pas disponibles. Elles sont estimées à partir d'un échantillon d'apprentissage dans lequel la classe associée à chaque observation est connue. Plusieurs méthodes d'estimation pourront être employées, parmi lesquelles les moyennes empiriques et matrices de covariance empiriques sont les plus répandues. Cependant, l'utilisation de ces techniques n'est pas toujours possible ou pourrait conduire à des performances médiocres. Ceci devient le cas si le nombre de variables et la taille de l'échantillon d'apprentissage sont du même ordre de grandeur. Dans ce travail, nous proposons une nouvelle technique d'estimation des matrices de covariance adaptée à des modèles de variance hétérogènes. Dans ces modèles, la matrice de covariance est supposée être une perturbation de rang fini d'une matrice proportionnelle à l'identité. L'estimateur proposé suit la même structure, avec une perturbation dont les valeurs propres associées sont des paramètres qui seront optimisés afin de maximiser les performances en classification. Curieusement, ces paramètres peuvent être optimisés facilement en utilisant uniquement l'échantillon d'apprentissage et sans recours aux méthodes de validation croisée. Des simulations extensives ont été effectuées pour tester les performances de la méthode proposée sur des données synthétiques et réelles et montrent qu'elle devance largement les techniques d'analyse discriminante classiques.

Abstract – Linear discriminant analysis is a classification technique widely used in the field of statistical learning. Under the Gaussian assumption, it corresponds to the optimal classification method in the sense of the maximum likelihood principle, when the classes have common covariance matrices, prefectly known as well as their respective means. In practice, these quantities are not available. They need to be estimated based on a set of training data for which the label of each observation is perfectly known. Several estimation methods can be used, among them, sample means and sample covariance matrices are the most frequently employed. However, the use of these techniques is not always possible or may lead to low performances. This is for instance the case when the number of features and the sample size are commensurable. In this work, we propose a novel estimation technique for the covariance matrix that is suitable to the cases of spiked covariance models. In these models, the covariance matrix is supposed to be a finite rank perturbation of a scaled identity. The proposed estimator follows the same structure, the largest eigenvalues of which are some parameters to be optimized so as to minimize the classification error rate. Interestingly, the optimal parameters admit a closed-form expression and can be set based on the sole knowledge of the training data, which avoids the need for using cross-validation approaches. Extensive simulations have been carried out to test the performance of the proposed method on a set of synthetic and real data, and were shown that it compares favorably with respect to classical discriminant analysis techniques.

1 Introduction

L'analyse discriminante linéaire est l'une des méthodes de référence utilisée dans le domaine de l'apprentissage statistique [1]. Sa popularité provient du fait qu'elle correspond à la méthode de classification qui présente l'erreur de classification minimale dans le cas où les données suivent un mélange gaussian avec une matrice de covariance commune [2].

Un obstacle majeur devant l'utilisation de l'analyse discriminante linéaire est lié au fait qu'elle se base sur la connaissance de l'inverse de la matrice de covariance et des vecteurs moyennes. Une technique très répandue en pratique consiste à concevoir des estimateurs de ces quantités à partir d'une base de donnée d'apprentissage pour laquelle la classe associée à chaque observation est connue. Parmi les estimateurs les plus utilisées sont les estimateurs de covariance empirique. Si le nombre d'échantillons dans la base d'apprentissage est suffisamment grand, ces estimateurs possèdent de bonnes qualités et donc les performances de la méthode de classification ne seront pas significativement altérées. Ceci n'est cependant pas la situation dans laquelle la taille de l'échantillon est du même ordre de grandeur ou même plus petite que la dimension des observations, comme envisagé dans les applications de traitement de données massives, de plus en plus en vogue aujourd'hui. Pour résoudre ce problème, il a été proposé de régulariser la matrice de covariance empirique par une matrice identité. Ceci permet de garantir qu'elle n'est pas singulière et donc son inverse peut être utilisée dans le score associé à l'analyse discriminante linéaire. Se pose alors la question du choix du coefficient de régularisation. Dans des travaux parus récemment, il a été proposé de sélectionner le coefficient de régularisation de telle façon à minimiser une approximation asymptotique de l'erreur de classification [3–5]. Cette approche, qui par ailleurs

constitue une alternative à la méthode de validation croisée, admet deux inconvénients majeurs. D'abord, la valeur optimale du coefficient de régularisation n'a pas de forme explicite, et donc ne peut être déterminée facilement. Enfin, cette approche ne permet pas d'exploiter une connaissance a priori portant sur la structure de la matrice de covariance. Ce papier propose une nouvelle méthode basée sur l'analyse discriminante linéaire qui s'applique dans les cas où la matrice de covariance est une perturbation de rang fini d'une matrice proportionnelle à l'identité. Ce modèle est présent dans les applications de traitement des signaux EEG [6], de détection [7] et en économétrie [8]. Plus spécifiquement, nous proposons d'utiliser un estimateur paramétré de la matrice de covariance qui respecte cette structure et dont les plus grandes valeurs propres sont les paramètres à optimiser. En utilisant des outils de la théorie des grandes matrices aléatoires, nous calculons une approximation asymptotique de l'erreur de classification. Ce résultat servira par la suite à obtenir les paramètres optimaux qui minimisent l'erreur de classification. Nous démontrons que ces paramètres admettent une forme explicite et sont faciles à calculer. La comparaison avec la méthode d'analyse linéaire discriminante classique montre que la méthode proposée offre des gains considérables, tant sur des données réelles que synthétiques. Le reste du papier est organisé comme suit. Nous rappelons en premier lieu les techniques d'analyse discriminante les plus utilisées. En second lieu, nous présentons la méthode proposée, et nous évaluons ses performances par comparaison avec les méthodes classiques.

2 Analyse discriminante linéaire

L'analyse discriminante linéaire est une méthode de classification supervisée dans laquelle un échantillon d'apprentissage est utilisé pour construire la régle de classification. Nous traitons le cas d'une classification binaire, où nous disposons de n observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ de taille $p \times 1$, composées de n_1 observations dans la classe $\mathcal{C}_1$ et n_0 observations dans la classe $\mathcal{C}_0$. Nous supposons que les observations de la classe $\mathcal{C}_k$, $k = 0, 1$ suivent la loi gaussienne de moyenne $\boldsymbol{\mu}_k$ et covariance $\boldsymbol{\Sigma}$. Ces quantités ne sont pas au préalable connues en pratique, elles sont estimées dans la plupart des cas par leurs estimateurs empiriques donnés par :

$$\overline{\mathbf{x}}_k = \frac{1}{n_k} \sum_{\ell \in \mathcal{T}_k} \mathbf{x}_\ell, \;\; k = 0, 1$$

$$\hat{\boldsymbol{\Sigma}} = \frac{(n_0 - 1)\hat{\boldsymbol{\Sigma}}_0 + (n_1 - 1)\hat{\boldsymbol{\Sigma}}_1}{n - 2}$$

où $\mathcal{T}_k$ désigne l'ensemble des indices des observations de classe $\mathcal{C}_k$ et

$$\hat{\boldsymbol{\Sigma}}_k = \frac{1}{n_k - 1} \sum_{\ell \in \mathcal{T}_k} (\mathbf{x}_\ell - \overline{\mathbf{x}}_\ell)(\mathbf{x}_\ell - \overline{\mathbf{x}}_\ell)^T, \;\; k = 0, 1.$$

est la matrice de covariance empirique associée à la classe $\mathcal{C}_k$. L'analyse discriminante linéaire se base sur le score discriminant donné par :

$$W^{\mathrm{LDA}}(\mathbf{x}) = (\mathbf{x} - \frac{\overline{\mathbf{x}}_0 + \overline{\mathbf{x}}_1}{2})^T \hat{\boldsymbol{\Sigma}}^{-1} (\overline{\mathbf{x}}_0 - \overline{\mathbf{x}}_1) - \log \frac{\pi_1}{\pi_0} \quad (1)$$

où $\pi_k, k = 0, 1$ est la probabilité a priori d'appartenance à la classe $\mathcal{C}_k$ supposée ici connue. Ce score, qui découle de l'application du principe de maximum de vraisemblance, est obtenu en remplaçant les matrices de covariances et les moyennes par leurs estimateurs empiriques. Il est ensuite utilisé pour prédire la classe d'une observation de test $\mathbf{x}$ en lui associant la classe $\mathcal{C}_0$ si $W^{\mathrm{LDA}}(\mathbf{x}) > 0$ et la classe $\mathcal{C}_1$ sinon. Si la taille des observations p est plus grande que la taille de l'échantillon n, la matrice de covariance empirique devient singulière, et donc ne peut être utilisée dans (1). Pour résoudre ce problème, la méthode la plus couramment utilisée consiste à régulariser la matrice de covariance de telle façon que toutes ses valeurs propres sont shiftées loin du zéro [2]. Ce faisant, le score discriminant devient :

$$W^{\mathrm{R-LDA}} = (\mathbf{x} - \frac{\overline{\mathbf{x}}_0 + \overline{\mathbf{x}}_1}{2})^T \mathbf{H} (\overline{\mathbf{x}}_0 - \overline{\mathbf{x}}_1) - \log \frac{\pi_1}{\pi_0} \quad (2)$$

où

$$\mathbf{H} = \left(\mathbf{I}_p + \gamma \hat{\boldsymbol{\Sigma}} \right)^{-1}, \;\; \gamma > 0$$

et γ désigne le coefficient de régularisation. La technique de classification correspondante est dénommée analyse discriminante régularisée en référence au facteur de régularisation utilisé. En conditionnant sur l'échantillon d'apprentissage, le score associé à (2) suit une loi gaussienne, étant une fonction linéaire du vecteur test $\mathbf{x}$. En conséquence, l'erreur de mauvaise classification conditionnelle de l'analyse discriminante linéaire associée à la classe $\mathcal{C}_k$ est donnée par :

$$\epsilon_k^{\mathrm{LDA}} = \Phi \left(\frac{(-1)^{k+1} G(\boldsymbol{\mu}_k, \overline{\mathbf{x}}_0, \overline{\mathbf{x}}_1, \hat{\boldsymbol{\Sigma}})}{\sqrt{D(\boldsymbol{\mu}_k, \overline{\mathbf{x}}_0, \overline{\mathbf{x}}_1, \hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})}} \right) \quad (3)$$

où $\Phi(.)$ est la fonction de répartition d'une variable normale standard et :

$$G(\boldsymbol{\mu}_k, \overline{\mathbf{x}}_0, \overline{\mathbf{x}}_1, \hat{\boldsymbol{\Sigma}}) = \left(\boldsymbol{\mu}_k - \frac{\overline{\mathbf{x}}_0 + \overline{\mathbf{x}}_1}{2} \right) \hat{\boldsymbol{\Sigma}}^{-1} (\overline{\mathbf{x}}_0 - \overline{\mathbf{x}}_1) \quad (4)$$

$$D(\boldsymbol{\mu}_k, \overline{\mathbf{x}}_0, \overline{\mathbf{x}}_1, \hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = (\overline{\mathbf{x}}_0 - \overline{\mathbf{x}}_1)^T \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \hat{\boldsymbol{\Sigma}}^{-1} (\overline{\mathbf{x}}_0 - \overline{\mathbf{x}}_1) \quad (5)$$

L'erreur de classification totale s'écrit alors :

$$\epsilon^{\mathrm{LDA}} = \pi_0 \epsilon_0^{\mathrm{LDA}} + \pi_1 \epsilon_1^{\mathrm{LDA}}. \quad (6)$$

De même, l'erreur de classification totale de l'analyse discriminante linéaire régularisée admet la même écriture avec la matrice $\hat{\boldsymbol{\Sigma}}^{-1}$ remplacée par $\mathbf{H}$. Cette erreur conditionnelle ne permet pas de révéler l'impact de la matrice de covariance et des moyennes théorique sur les performances en matière de classification. Pour cela, de nouvelles méthodes basées sur l'utilisation des outils asymptotiques de la théorie des matrices aléatoires ont été utilisés pour donner des approximations précises des quantités (4) et (5) qui ne dépendent que des quantités théoriques $\boldsymbol{\Sigma}$ et $\boldsymbol{\mu}_k, k = 0, 1$.

L'analyse discriminante régularisée telle que décrite ci-dessus, est présentée dans les travaux d'apprentissage statistique comme une méthode de référence utilisée dans les comparaisons avec d'autres méthodes de classification. Cependant, elle possède deux inconvenients majeurs. Premièrement, il a été observé

qu'une attention particulière devra être accordée au choix de facteur de régularisation, un choix arbitraire pouvant conduire à des pertes significatives en matière de performance de classification. Il est courant que ce choix soit réalisé en testant les performances de quelques valeurs candidates par validation croisée. Cette approche a non seulement une complexité élevée mais aussi ne permet pas une compréhension qualitative du rôle joué par le facteur de régularisation dans les performances de classification. Deuxièmement, l'utilisation de la matrice de covariance empirique avec régularisation ou non ne permet pas d'exploiter une certaine connaissance a priori portant sur la structure de la matrice de covariance. Il est en effet courant selon le type d'application concerné que l'on a une idée sur la structure de la matrice de covariance. Ceci est le cas par exemple des applications de traitement des signaux EEG, ou les problèmes de détection, dans lesquelles la matrice de covariance $\mathbf{\Sigma}$ pourra s'écrire comme : $\mathbf{\Sigma} = \sigma^2\mathbf{I}_p + \sigma^2\sum_{j=1}^{r}\lambda_j\mathbf{v}_j\mathbf{v}_j^T$ où $\sigma^2 > 0$, $\lambda_1, \geq \ldots, \lambda_r > 0$ et $\mathbf{v}_1, \ldots, \mathbf{v}_r$ forment une famille orthonormale de vecteurs. Ce travail propose une nouvelle méthode d'analyse discriminante qui exploite cette structure particulière.

3 Méthode proposée

Par souci de simplicité nous supposons que les quantités r et σ^2 sont parfaitement connues. En pratique, elles peuvent être estimées à l'aide de plusieurs méthodes existantes dans la littérature. Nous reférons le lecteur aux travaux suivants [9,10]. En notant que :

$$\mathbf{\Sigma}^{-1} = \frac{1}{\sigma^2}\left[\mathbf{I}_p - \sum_{j=1}^{r}\frac{\lambda_j}{1+\lambda_j}\mathbf{v}_j\mathbf{v}_j^T\right]$$

il paraît judicieux de chercher des estimateurs de $\mathbf{\Sigma}^{-1}$ de la forme :

$$\hat{\mathbf{C}}^{-1} = \frac{1}{\sigma^2}\left[\mathbf{I}_p + \sum_{j=1}^{r}\omega_j\mathbf{u}_j\mathbf{u}_j^T\right] \tag{7}$$

où $\mathbf{u}_1, \ldots, \mathbf{u}_r$ sont les r vecteurs propres associés aux r valeurs propres les plus grandes de $\hat{\mathbf{\Sigma}}$ et $\{\omega_j\}_{j=1}^{r}$ sont des paramètres à optimiser dans l'intervalle $(-1, 0)$. En remplaçant $\hat{\mathbf{\Sigma}}$ par $\hat{\mathbf{C}}^{-1}$ dans (1), on obtient le score discriminant suivant :

$$W^{\text{Imp-LDA}} = \left(\mathbf{x} - \frac{\overline{\mathbf{x}}_0 + \overline{\mathbf{x}}_1}{2}\right)\hat{\mathbf{C}}^{-1}(\overline{\mathbf{x}}_0 - \overline{\mathbf{x}}_1) - \log\frac{\pi_1}{\pi_0}$$

Il est à noter que l'estimateur $\hat{\mathbf{C}}^{-1}$ peut être vu comme une instance d'une classe plus large d'estimateurs de matrice de covariance qui sont de la forme $\sum_{j=1}^{p}\eta_j\mathbf{u}_j\mathbf{u}_j^T$, où $\{\mathbf{u}_j\}_{j=1}^{p}$ sont les vecteurs propres de la matrice de covariance empirique et leurs valeurs propres $\eta_j, j = 1\ldots, p$ prennent la forme de fonctions de rétrécessisement qui doivent être conçues dans le but de répondre à un objectif donné. En particulier, $\hat{\mathbf{C}}^{-1}$ correspond à choisir $\eta_{r+1} = \cdots = \eta_p = \frac{1}{\sigma^2}$ et $\omega_j = \sigma^2\eta_j - 1, j = 1, \ldots, r$. Compte tenu de l'application en notre disposition, il fait naturellement sens de choisir les coefficients $\omega_j, j = 1\ldots, r$ de telle façon à minimiser l'erreur de classification globale. Plus formellement, le vecteur $\boldsymbol{\omega}^\star = [\omega_1, \ldots, \omega_r]$ est tel que : $\boldsymbol{\omega}^\star = \arg\min_{\boldsymbol{\omega}}\mathbb{E}\left[\epsilon^{\text{Imp-LDA}}\right]$ où $\epsilon^{\text{Imp-LDA}}$ est obtenu en remplaçant $\hat{\mathbf{\Sigma}}^{-1}$ par $\hat{\mathbf{C}}^{-1}$ dans (6). Une analyse approfondie de l'erreur de classification est donc requise mais ceci ne peut être réalisé par des méthodes analytiques exactes. Pour cette raison, nous faisons appel à des outils avancés de la théorie des matrices aléatoires, qui permettront de fournir une bonne approximation de l'erreur de classification globale dans le cas où n et p sont larges et du même ordre de grandeur. Plus spécifiquement, nous considérons l'hypothèse suivante, sous laquelle nous démontrons le Théorème 1.

Hypothèse 1.
— $n, p \to \infty$ *avec* $c = \frac{p}{n}$.
— $\frac{n_1}{n} = \pi_1$ *et* $\frac{n_0}{n} = \pi_0$.
— r *est fixe et* $\lambda_1 > \cdots > \lambda_r > \sqrt{c}$.

Théorème 1. *En accord avec les notations précédentes, sous l'hypothèse 1, nous avons :*

$$G(\boldsymbol{\mu}_i, \overline{\mathbf{x}}_0, \overline{\mathbf{x}}_1, \hat{\mathbf{C}}) - \frac{1}{2}\left[\frac{(-1)^i\|\boldsymbol{\mu}\|^2}{\sigma^2}\overline{G}(\boldsymbol{\omega}) - \frac{p}{n_0} + \frac{p}{n_1}\right] \xrightarrow{a.s.} 0,$$

$$D(\boldsymbol{\mu}_i, \overline{\mathbf{x}}_0, \overline{\mathbf{x}}_1, \hat{\mathbf{C}}, \mathbf{\Sigma}) - \left[\frac{\|\boldsymbol{\mu}\|^2}{\sigma^2}\overline{D}(\boldsymbol{\omega}) + \frac{p}{n_0} + \frac{p}{n_1}\right] \xrightarrow{a.s.} 0,$$

où $\overline{G}(\boldsymbol{\omega}) = 1 + \sum_{j=1}^{r} a_j b_j w_j$, $\overline{D}(\boldsymbol{\omega}) = 1 + \sum_{j=1}^{r}\left[a_j b_j(1+\lambda_j a_j)\omega_j^2\right]$ $+\sum_{j=1}^{r}\left[\lambda_j b_j + 2a_j b_j(\lambda_j+1)\omega_j\right]$, $\boldsymbol{\mu} = \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1$, $a_j = \frac{1-c/\lambda_j^2}{1+c/\lambda_j}$ *et* $b_j = \frac{\boldsymbol{\mu}^T\mathbf{v}_j\mathbf{v}_j^T\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|^2}$, $j = 1, \cdots, r$.

Le résultat du Théorème 1 permet de donner une approximation de l'erreur de classification globale. En particulier, nous avons par application du théorème de convergence dominé :

$$\mathbb{E}\left[\epsilon^{\text{Imp-LDA}}\right] - \bar{\epsilon}^{\text{Imp-LDA}} \longrightarrow 0, \text{ avec}$$

$$\bar{\epsilon}^{\text{Imp-LDA}} = \pi_0\Phi\left(\frac{-\alpha(\overline{G}(\boldsymbol{\omega}) - \eta)}{\sqrt{\overline{D}(\boldsymbol{\omega}) + \theta}}\right) + \pi_1\Phi\left(\frac{-\alpha(\overline{G}(\boldsymbol{\omega}) + \eta)}{\sqrt{\overline{D}(\boldsymbol{\omega}) + \theta}}\right)$$

avec $\alpha = \frac{\|\boldsymbol{\mu}\|}{2\sigma}$, $\eta = \frac{\sigma^2}{\|\boldsymbol{\mu}\|^2}[c/\pi_0 - c/\pi_1 + 2\log\frac{\pi_1}{\pi_0}]$ et $\theta = \frac{\sigma^2}{\|\boldsymbol{\mu}\|^2}[p/n_0 + p/n_1]$. Cette approximation est maintenant exploitée afin de déterminer les coefficients $\omega_j^\star, j = 1\ldots, r$ qui la minimisent. En ce faisant, nous avons le résultat suivant :

Théorème 2. *Sous l'hypothèse 1, les paramètres* $\omega_j^\star, j = 1\ldots, r$ *qui minimisent* $\bar{\epsilon}$ *sont donnés par :*

$$\omega_j^\star = \frac{u^\star}{\beta}\frac{\gamma_j}{\alpha_j} - \beta_j, \quad j = 1, \cdots, r, \quad \text{où}$$

$$\alpha_j = \lambda_j a_j^2 b_j + a_j b_j, \ \beta_j = \frac{\lambda_j + 1}{\lambda_j a_j + 1}, \ \gamma_j = a_j b_j, \ j = 1, ..., r,$$

$\beta = \sqrt{\sum_{j=1}^{r}\gamma_j^2/\alpha_j}$, *et* $u^\star$ *est l'unique solution positive de l'équation* $h(u) = 0$, *avec*

$$h(u) = \pi_0(\beta b - d_0 u)e^{\frac{-\alpha^2}{2}\frac{(\beta u + d_0)^2}{u^2+b}} + \pi_1(\beta b - d_1 u)e^{\frac{-\alpha^2}{2}\frac{(\beta u + d_1)^2}{u^2+b}}, \tag{8}$$

$$b=1+\theta+\sum_{j=1}^{r}\left[\lambda_j b_j-\frac{(\lambda_j a_j b_j+a_j b_j)^2}{\lambda_j a_j^2 b_j+a_j b_j}\right],$$

$$d_0=1-\eta-\sum_{j=1}^{r}\frac{\lambda_j a_j b_j+a_j b_j}{\lambda_j a_j+1},\ d_1=1+\eta-\sum_{j=1}^{r}\frac{\lambda_j a_j b_j+a_j b_j}{\lambda_j a_j+1}.$$

Il est à noter que lorsque les classes sont équiprobables ($\pi_0 = \pi_1 = 0.5$), $u^\star$ est donné en forme explicite par $u^\star = \frac{\beta b}{d}$ avec $d = 1 - \sum_{j=1}^{r} \frac{\lambda_j a_j b_j + a_j b_j}{\lambda_j a_j + 1}$. En pratique, il n'est toujours pas possible d'utiliser les valeurs optimales $\omega_j^\star, j = 1 \ldots, r$ car elles dépendent de la matrice de covariance à travers, $\lambda_j, j = 1, \ldots, r$ et b_j, lesquelles ne sont pas observables. Pour résoudre ce problème, des estimateurs consistents de ces paramètres doivent être calculés. Ceci constitue l'objectif du résultat suivant :

Proposition 1. *Sous l'hypothèse 1, nous avons.* $|\alpha - \hat{\alpha}| \xrightarrow{a.s.} 0$, $|\lambda_j - \hat{\lambda}_j| \xrightarrow{a.s.} 0$, et $|b_j - \hat{b}_j| \xrightarrow{a.s.} 0$, avec

$$\hat{\alpha} = \frac{\|\hat{\boldsymbol{\mu}}\|^2}{\sigma^2} - c_1 - c_0$$

$$\hat{\lambda}_j = \frac{s_j/\sigma^2 + 1 - c + \sqrt{(s_j/\sigma^2 + 1 - c)^2 - 4s_j/\sigma^2}}{2} - 1,$$

$$\hat{b}_j = \frac{1 + c/\hat{\lambda}_j}{1 - c/\hat{\lambda}_j^2} \frac{\hat{\boldsymbol{\mu}}^T \mathbf{u}_j \mathbf{u}_j^T \hat{\boldsymbol{\mu}}}{\|\hat{\boldsymbol{\mu}}\|^2 - c_1\sigma^2 - c_0\sigma^2}, \quad j = 1, \cdots, r.$$

avec $\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}}_0 - \overline{\mathbf{x}}_1$ *et* $s_j = 1, \ldots, r$ *désigne la* j*-ème plus grande valeur propre de la matrice de covariance emprique.*

4 Résultats des simulations

Nous présentons ici les résultats des performances de la méthode de classification proposée par comparaison avec la méthode discriminante régularisée qu'on désigne par R-LDA. Nos simulations sont effectuées sur des données synthétiques et des données réelles. Faute de place, nous nous contentons de présenter les résultats de celles appliquées sur des données réelles. Nous considérons des données qui viennent de la base de données USPS disponible sur le site `http://www.csie.ntu.edu.tw/cjlin/libsvmtools`. Elle est composée de 9298 d'images de chiffres écrits à la main où chaque image est représentée sous la forme d'un vecteur d'observation de taille $p = 256$. Nous considérons la classification des images affichant des chiffres 4 et 9 écrits à la main lorsque $\pi_0 = \pi_1 = 0.5$. Nous représentons dans la Figure 1 l'erreur de classification totale versus le nombre d'observations n. Nous notons un gain significatif de la méthode proposée par rapport à la méthode classique R-LDA.

5 Conclusion

Ce travail propose une nouvelle méthode basée sur l'analyse discriminante linéaire. Cette méthode est adaptée aux scénarios où la matrice de covariance suit la structure particulière d'un modèle de variance hétérogène. La technique proposée est basée sur l'utilisation d'un estimateur paramétrique de la matrice de covariance qui suit la même structure. Ses paramètres sont optimisés afin de maximiser une approximation asymptotique de l'erreur de classification. Les résultats des simulations sur des données réelles sont très encourageants et montrent un gain important par rapport aux techniques d'analyse discriminante classiques.

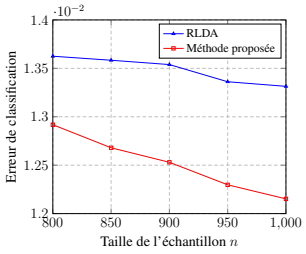

FIGURE 1 – Erreur de classification versus la taille de l'échantillon n. Comparaison entre la méthode proposée et l'analyse discriminante régularisée.

Références

[1] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, 1936.

[2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2001.

[3] A. Zollanvari and E. R. Dougherty, "Generalized consistent error estimator of linear discriminant analysis," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2804–2814, Jun. 2015.

[4] Daniyar Bakirov, Alex Pappachen James, and Amin Zollanvari, "An efficient method to estimate the optimum regularization parameter in RLDA," *Bioinformatics*, vol. 32, no. 22, pp. 3461–3468, 2016.

[5] Khalil Elkhalil, Abla Kammoun, Romain Couillet, Tareq Y. Al-Naffouri, and Mohamed-Slim Alouini, "A Large Dimensional Study of Regularized Discriminant Analysis Classifiers," *https ://arxiv.org/abs/1711.00382*, 2017.

[6] G. Huang, G. Liu, J. Meng, D. Zhang, and X. Zhu, "Model based generalization analysis of common spatial pattern in brain computer interfaces," *Cogn. Neurodyn*, vol. 4, pp. 217223, 2010.

[7] S. Kritchman and B. Nadler, "Non-Parametric Detection of the Number of Signals : Hypothesis Testing and Random Matrix Theory," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, Oct. 2009.

[8] J. Fan, Y. Fan, and J. Lv, "High dimensional covariance matrix estimation using a factor model," *Journal of Econometrics*, vol. 147, no. 1, pp. 186–197, 2008.

[9] M. O. Ulfarsson and V. Solo, "Dimension estimation in noisy pca with sure and random matrix theory," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 5804–5816, Dec. 2008.

[10] D. Passemier, Z. Li, and J. Yao, "On estimation of the noise variance in high dimensional probabilistic principal component analysis," *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, vol. 79, no. 1, pp. 51–67, 2017.