2.4 Séance de TP sur la méthode probabiliste

Remarques générales
Cette séance de travaux pratiques (TP) est l'occasion de mettre en œuvre les théories étudiées en cours. Il est donc indispensable d'avoir une bonne connaissance du cours pour analyser les résultats des expériences menées pendant la séance. Le sujet abordé pour ce deuxième TP est la discrimination dans un problème à C classes.

Les étudiants doivent construire leur raisonnement en s'appuyant sur une démarche expérimentale. Pour les aider dans cette démarche, ils sont encouragés à poser des questions aux enseignants pendant la séance. Ce TP donne lieu à un compte rendu écrit d'une page A4 (recto uniquement, taille des polices 11) qui est remis avant le jour suivant la séance de TP au format pdf.

Le rôle de ce compte rendu est de décrire les quelques points qui vous paraissent essentiels et non de faire une liste exhaustive des points abordés en TP. Le compte rendu doit aussi poser des questions destinées à alimenter une discussion qui aura lieu lors du prochain cours. Chaque groupe travaille à son rythme et rédige son compte rendu en autonomie.

Contexte de l'étude

Dans cette étude, les vecteurs de caractéristiques $\boldsymbol{x} \in \mathbb{R}^N$ sont distribués suivant des lois normales :

$$f(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Gamma}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Gamma}|^{1/2}} \exp\left[-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right] \tag{43}$$

où $\boldsymbol{\mu}$ est la moyenne de $\boldsymbol{x}$, $\boldsymbol{\Gamma}$ la matrice de covariance de $\boldsymbol{x}$ et $^{-1}$ représente l'opérateur inverse si $\det(\boldsymbol{\Gamma}) > 0$ et la pseudo-inverse sinon. De plus, dans le problème de discrimination considéré, à tout vecteur $\boldsymbol{x}$, on associe un numéro de classe $d \in \{1,\ 2,\ ...,\ C\}$. Tous les vecteurs $\boldsymbol{x} \in \mathbb{R}^N$ associés à une même classe sont identiquement distribués. On a donc C vecteurs de moyenne $(\boldsymbol{\mu}_c)_{c\in\{1,\ 2,\ ...,\ C\}}$ et C matrices de covariance $(\boldsymbol{\Gamma}_c)_{c\in\{1,\ 2,\ ...,\ C\}}$ pour caractériser les différentes classes. On analyse les performances de 3 discriminateurs (voir ci-dessous).

Discriminateur avec moyennes et matrices de covariance connues

$$\widehat{d}_{\mu,\Gamma \text{ connus}}(\boldsymbol{x}) = \arg \max_{c\in\{1,\dots,C\}} \left(f(\boldsymbol{x}, (\boldsymbol{\mu}_c, \boldsymbol{\Gamma}_c))\right) \tag{44}$$

Discriminateur linéaire

$$\widehat{d}_{lin}(\boldsymbol{x}) = \arg \max_{c\in\{1,\dots,C\}} \left(f(\boldsymbol{x}, (\widehat{\boldsymbol{\mu}}_c, \widehat{\boldsymbol{\Gamma}}_{lin}))\right) \tag{45}$$

où les paramètres $\widehat{\boldsymbol{\mu}}_c$ et $\widehat{\boldsymbol{\Gamma}}_{lin}$ sont estimés au sens du maximum de vraisemblance en supposant que les classes partagent la même matrice de covariance : $\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_2 = ... = \boldsymbol{\Gamma}_c$.

Discriminateur quadratique

$$\widehat{d}_{quad}(\boldsymbol{x}) = \arg \max_{c\in\{1,\dots,C\}} \left(f(\boldsymbol{x}, (\widehat{\boldsymbol{\mu}}_c, \widehat{\boldsymbol{\Gamma}}_c))\right) \tag{46}$$

où contrairement au discriminateur linéaire, on ne suppose pas que : $\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_2 = ... = \boldsymbol{\Gamma}_c$, mais on contraire on estime les matrices de covariance $\widehat{\boldsymbol{\Gamma}}_c$ pour chacune des classes.

Apprentissage
On dispose d'une base d'apprentissage $\left(\boldsymbol{x}^{(p)}\right)_p$ de taille P_{app} où les classes sont équiréparties. Soit $I_{app,c}$ les indices de la base d'apprentissage correspondant à la classe c, on a :

$$\begin{array}{l} \widehat{\boldsymbol{\mu}}_c = \frac{C}{P_{app}} \sum_{p\in I_{app,c}} \boldsymbol{x}^{(p)} \\ \widehat{\boldsymbol{\Gamma}}_c = \frac{C}{P_{app}} \sum_{p\in I_{app,c}} \left(\boldsymbol{x}^{(p)} - \widehat{\boldsymbol{\mu}}_c\right)\left(\boldsymbol{x}^{(p)} - \widehat{\boldsymbol{\mu}}_c\right)^T \\ \widehat{\boldsymbol{\Gamma}}_{lin} = \frac{1}{C} \sum_{c=1}^{C} \widehat{\boldsymbol{\Gamma}}_c \end{array} \tag{47}$$

Questions à traiter

2.4.1 Analyse des surfaces discriminantes dans $\mathbb{R}^2$
Dans cette partie, on analyse des bases de données simulées dans un problème à 2 classes pour lequel $\boldsymbol{x} \in \mathbb{R}^2$. En deux dimensions, on visualise les éléments de la base dans différentes situations : matrices de covariances identiques ou pas (i.e. $\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_2$ ou $\boldsymbol{\Gamma}_1 \neq \boldsymbol{\Gamma}_2$), signaux centrés (i.e. $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = 0$) ou pas.

1.1/ **Cas particulier où $\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_2$ et $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$**
Exécutez une dizaine de fois le script `main.m` pour générer différents exemples de base de données avec chacune $P = 1000$ éléments. Commentez les résultats obtenus. Tracez au tableau l'allure des densités de probabilité en 2 dimensions (2D) et en déduire la forme d'une frontière simple permettant de séparer les deux classes.

1.2/ **Cas particulier où $\boldsymbol{\Gamma}_1 \neq \boldsymbol{\Gamma}_2$ et $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = 0$**
Modifiez le script à la ligne 14 : `choix_centre=1;` pour choisir des cas où les vecteurs sont centrés $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = 0$ ainsi qu'à la ligne 18 : `choix_meme_matrice_de_covariance=0;` pour que les matrices de covariances sont différentes $\boldsymbol{\Gamma}_1 \neq \boldsymbol{\Gamma}_2$. Commentez les résultats obtenus. En particulier, tracez au tableau l'allure des densités de probabilité en 2 dimensions (2D) et en déduire la forme d'une frontière simple pour séparer les deux classes.

1.3/ **Cas particulier où $\boldsymbol{\Gamma}_1 \neq \boldsymbol{\Gamma}_2$ et $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$**
Modifiez à nouveau le script à la ligne 14 : `choix_centre=0;` pour considérer le cas où les vecteurs ne sont pas centrés $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ et les matrices de covariances sont différentes $\boldsymbol{\Gamma}_1 \neq \boldsymbol{\Gamma}_2$. Commentez les résultats obtenus. En particulier, tracez au tableau l'allure des densités de probabilité en 2 dimensions (2D) et en déduire la forme d'une frontière simple pour séparer les deux classes.

1.4/ **Surface Discriminantes**
Modifiez le programme `main` avec `trace_surf_discri=1` et relancez le programme pour tracer la surface discriminante obtenue par le discriminateur avec $\boldsymbol{\mu}$ et $\boldsymbol{\Gamma}$ connus. En relisant vos notes de cours, répondez aux questions suivantes. Quel est le critère optimisé par le discriminateur avec $\boldsymbol{\mu}$ et $\boldsymbol{\Gamma}$ connus ? Quel est le lien entre les densités de probabilités et la surface discriminante obtenue ? Rappelez l'équation suivie par la surface discriminante. Ce résultat remet-il en cause vos observations des questions précédentes ? Validez avec un encadrant.

2.4.2 Comparaison des performances de 3 discriminateurs
Dans cette partie, on considère plusieurs exemples simulés en gaussien, et pour chacun on quantifie les performances des 3 discriminateurs considérés avec une base de généralisation de taille P_{gen}.

2.1/ On considère un problème à 3 classes. Exécutez le programme avec `question=2`. Vous obtenez le taux de généralisation τ_g du discriminateur avec $\boldsymbol{\mu}$ et $\boldsymbol{\Gamma}$ connus. Comment obtenir la barre d'erreur sur l'estimation du taux de généralisation ? Vérifiez votre barre d'erreur en relançant plusieurs fois le programme. Validez avec un encadrant.

2.2/ Sur le même problème de discrimination, on compare les performances du discriminateur avec $\boldsymbol{\mu}$ et $\boldsymbol{\Gamma}$ connus avec le discriminateur dit "linéaire" (où $\boldsymbol{\mu}$ et $\boldsymbol{\Gamma}$ sont estimés, voir première page). Modifiez le programme avec `choix_discri=2`. Vous obtenez figure 1 une réalisation d'une base d'apprentissage avec $P_{app} = 100$ éléments. Est-ce que vous observez une perte de performance significative avec le discriminateur linéaire ? Pourquoi ? Validez avec un encadrant.

2.3/ On ajoute à l'analyse le discriminateur quadratique (`choix_discri=3`). Est-ce que vous observez des différences de performance significatives ? Pourquoi ? Comme on est en dimension 2, on pourra tracer les surfaces discriminantes de chaque discriminateur (`choix_trace_surf_discri=1`). Validez avec un encadrant.

2.4/ On considère un nouvel exemple. Modifiez le programme avec `choix_exemple=2`. Est-ce que vous observez des différences de performance significatives ? Pourquoi ? Validez avec un encadrant.

2.5/ Diriez vous que le dicton : "qui peut le plus peut le moins" s'applique à vos observations sur le "linéaire" et le "quadratique" ? Pourquoi ? Validez avec un encadrant.

2.6/ On considère un nouvel exemple pour lequel la dimension de l'espace est $N = 38$ (`choix_exemple=3`). On analyse l'évolution des performances en fonction de la taille de la base d'apprentissage (enlevez le `%` ligne 53 du programme). Commentez et interprétez les résultats obtenus. Est-il possible d'améliorer les résultats ? Validez avec un encadrant.