Aix*Marseille université

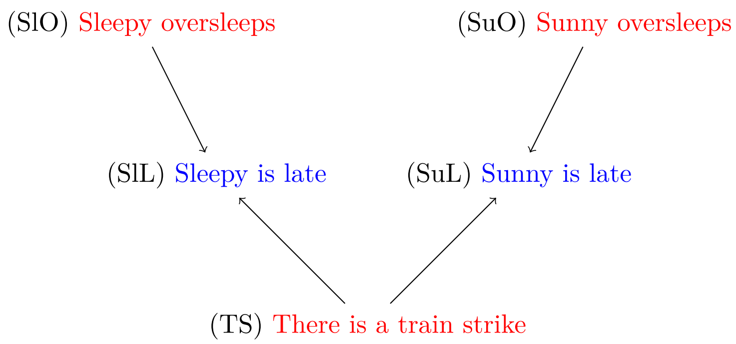
MASTER 2 INFORMATIQUE

PARCOURS IAAA, UE MIA

Année 2021-22

DM 3 : probabilités et statistique

Exercice 1 Calcul probabiliste.



Considérons le modèle graphique représenté ci-dessus. SlO, SuO, SlL, SuL and TS sont des variables aléatoire binaires prenant leurs valeurs dans $\{0, 1\}$. Dans cet exercice nous allons essayer de déterminer ce qui peut être inféré sur les variables latentes (en rouge) à partir de l'observation des variables observées (en bleu).

Nous faisons l'hypothèse que si au moins un des deux évènements "Sleepy oversleeps" et "There is a train strike" a lieu, alors 'Sleepy is late' a lieu également (avec probabilité 1). De façon similaire, si au moins un des deux évènements 'Sunny oversleeps' et 'There is a train strike' a lieu, alors 'Sunny is late' a lieu également (avec probabilité 1). Nous pouvons l'écrire plus formellement, de la manière suivante :

$$P(SlL = 1|SlO = a, TS = b) = a \vee b,$$

et :

$$P(SlL = 1|SlO = a, TS = b) = a \vee b,$$

pour tout a, b dans $\{0, 1\}$. Le symbole $\vee$ représente le connecteur logique *ou* (inclusif) de $\{0, 1\}$ vers $\{0, 1\}$.

On note $l = P(SlO = 1)$, $u = P(SuO = 1)$ et $t = P(TS = 1)$.

1. Donner la factorisation de $P(SlL, SuL, SlO, SuO, TS)$ d'après le modèle graphique représenté ci-dessus.
2. La distribution de probabilité $P(SlL, SuL, SlO, SuO, TS)$ est-elle entièrement déterminée si les valeurs de l, u et t sont données ?

3. Calculer $P(TS = 1|SlL = 1)$ en fonction de l, u et t.
4. Calculer $P(SlO = 1|SlL = 1)$ en fonction de l, u et t.
5. Calculer $P(TS = 1|SlL = 1, SuL = 1)$ en fonction de l, u et t.
6. Calculer $P(SlO = 1|SlL = 1, SuL = 1)$ en fonction de l, u et t.
7. Supposer à présent que $l = 0.5$, $t = 0.1$ et que lévenement 'Sleepy is late' a lieu. Quel évènement est alors le plus probable : 'There is a train strike' ou 'Sleepy overslept' ?
8. Même question si on suppose en plus que $u = 0.01$ et que l'évènement 'Sunny is late' est également observé.
9. Que se passe-t-il si on prend $l = 0.5$, $t = 0.1$ et $u = 0.2$?

Exercice 2 Analyse des propriétés basiques d'un estimateur.

Supposons qu'on observe un échantillon $x_1, x_2, \ldots, x_n \in (\mathbf{R}^d)^n$, i.i.d. de distribution P. Soit $d : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$ une fonction mesurant la 'dissimilarité' entre deux points de $\mathbf{R}^d$. On suppose que d est symmétrique, c'est à dire que $d(x_1, x_2) = d(x_2, x_1)$ pour tout choix de x_1, x_2. Nous cherchons à estimer :

$$\delta(P, d) := \mathbf{E}_{a,b\sim P\otimes P}[d(a, b)]$$

sur la base de $x_1, x_2, ..., x_n$ ($a, b \sim P \otimes P$ signifie que a et b sont deux échantillons tirés indépendamment de la loi P). On suppose que $\mathbf{E}_{a,b\sim P\otimes P}[d(a, b)^2] < +\infty$.

Considérons l'estimateur $\hat{\delta}(x_1, x_2, \ldots, x_n) := \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{j=i+1}^{n} d(x_i, x_j)$.

1. Quel est le biais de $\hat{\delta}$?
2. Quelle est la variance de $\hat{\delta}$? On l'exprimera en fonction de $\sigma_1 = \mathrm{Var}_{x_1\sim P}\mathbf{E}_{x_2\sim P}[d(x_1, x_2)]$ et $\sigma_2 = \mathrm{Var}_{(x_1,x_2)\sim P\otimes P}[d(x_1, x_2)]$.
3. Prouver l'inégalité de Markov : soit X est un variable aléatoire réelle positive, de moyenne finie μ et soit t un réel strictement positif, alors :

$$p(X \geq t) \leq \frac{\mu}{t}.$$

4. Utiliser l'inégalité de Markov pour prouver l'inégalité de Chebyshev : soit X est un variable aléatoire réelle positive, de moyenne finie μ et de variance finie σ^2 et soit t un réel strictement positif, alors :

$$p(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

5. Utiliser l'inégalité de Chebyshev et les résultats des deux premières questions pour montrer que $\hat{\delta}$ est un estimateur faiblement consistant de δ.

Exercice 3 Estimation par maximum de vraisemblance.

Supposons qu'on observe un échantillon i.i.d. $(x_1, x_2, \ldots, x_n) \in (\mathbf{R}^d)^n$ de loi gaussienne multivariée $\mathcal{N}(\mu^*, \Sigma^*)$, pour un vecteur $\mu \in \mathbf{R}^d$ et une matrice $\Sigma^* \in \mathbf{S}_d$, où $\mathbf{S}_d$ est l'ensemble des matrices à coefficients réels symétriques définies positives de taille d pas d. On cherche à estimer μ^* et Σ^* à partir de l'observation de $(x_1, x_2, \ldots, x_n)$.

On définit la *vraisemblance* d'un couple de paramètres $(\mu, \Sigma) \in \mathbf{R}^d \times \mathbf{S}_d$ comme :

$$\ell(\mu, \Sigma) := p(x_1, x_2, \ldots, x_n; \mu, \Sigma),$$

où p corresponds à la densité de probabilité de $x_1, x_2, \ldots, x_n$ (par rapport à la mesure de Lebesgue sur $(\mathbf{R}^d)^n$).

On considère l'estimateur du *maximum de vraisemblance* pour (μ^*, Σ^*) défini par

$$\hat{\mu}, \hat{\Sigma} \in \arg \max_{(\mu,\Sigma) \in \mathbf{R}^d \times \mathbf{S}_d} \ell(\mu, \Sigma).$$

1. Montrer que
$$\hat{\mu}, \hat{\Sigma} \in \arg \max_{(\mu,\Sigma) \in \mathbf{R}^d \times \mathbf{S}_d} \log(\ell(\mu, \Sigma)).$$
2. Donner une expression la plus simple possible pour $\log(\ell(\mu, \Sigma)))$ en utilisant la formule donnant la densité d'une loi gaussienne multivariée non dégénérée.
3. On suppose que $\log(\ell(\mu, \Sigma)))$ est de classe C^1 et admet un unique maximum sur $\mathbf{R}^d \times \mathbf{S}_d$ en un point où son gradient s'annule. Calculer une expression explicite pour $(\hat{\mu}, \hat{\Sigma})$ en fonction de $x_1, x_2, \ldots, x_n$.

Exercice 4 Calculs probabilistes pour l'algorithme EM appliqué à un mélange de gaussiennes.

Supposons qu'on observe un échantillon i.i.d. $(x_1, x_2, \ldots, x_n) \in (\mathbf{R}^d)^n$ tiré d'un mélange de K gaussiennes, non-dégénéré, c'est à dire d'une distribution de probabilité dont la densité (par rapport à la mesure de Lebesgue sur $\mathbf{R}^d$) est donnée par :

$$p(x; \pi_1, \pi_2, \ldots, \pi_K, \mu_1, \mu_2, \ldots, \mu_K, \Sigma_1, \Sigma_2, \ldots, \Sigma_K := \sum_{k=1}^{K} \pi_k \mathcal{N}(x; \mu_k, \Sigma_k),$$

où pour tout $k \in \{1, 2, \ldots, K\}$, π_k est un réel positif, $\mu_k \in \mathbf{R}^d$ et Σ_k est une matrice symétrique définie positive à coefficients réels de taille d par d, $\sum_{k=1}^{K} \pi_k = 1$ et $\mathcal{N}(.; \mu, \Sigma)$ est la densité d'une loi gaussienne multivariée de paramètres μ et Σ.

Supposons que nous avons accès pour tout i dans $\{1, 2, \ldots, n\}$ et pour tout k dans $\{1, 2, \ldots, K\}$, à la quantité :

$$r_{i,k} = \frac{\pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i; \mu_j, \Sigma_j)},$$

mais que nous ne connaissons pas la valeur des paramètres $\pi_1, \pi_2, \ldots, \pi_K, \mu_1, \mu_2, \ldots, \mu_K, \Sigma_1, \Sigma_2, \ldots, \Sigma_K$.

1. Donner une expression, la plus simple possible, pour le logarithme de la vraisemblance au vu de $x_1, x_2, \ldots x_n$ (définie dans l'exercice précédent) pour un choix de paramètres $\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_K$, $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K$, $\hat{\Sigma}_1, \hat{\Sigma}_2, \ldots, \hat{\Sigma}_K$.
2. Supposons que le logarithme de la vraisemblance est de classe $\mathcal{C}^1$ et admet un unique maximum sur son ensemble de définition en un point où son gradient s'annule. Donner une expression explicite pour l'estimateur du maximum de vraisemblance pour $\pi_1, \pi_2, \ldots, \pi_K, \mu_1, \mu_2, \ldots, \mu_K, \Sigma_1, \Sigma_2, \ldots, \Sigma_K$ en fonction de $x_1, x_2, \ldots, x_n$ et des $r_{i,k}$, $i \in \{1, 2, \ldots, n\}$, $k \in \{1, 2, \ldots, K\}$.

Exercice 5 Weyl's inequality.

Nous allons prouver un résultat qui peut être utile, par exemple, pour analyser les propriétés statistiques de l'estimateur usuel utilisé dans le cadre d'une analyse en composantes principales.

Considérons une matrice symmétrique M à coefficients réels et de taille d par d à laquelle on ajoute une perturbation, représentée par une autre matrice symétrique P à coefficients réels et de taille d par d. On cherche à trouver des conditions sous lesquelles on peut garantir que les valeurs propres de M et celles de $M' = M + P$ sont proches.

Pour toute matrice symmétrique A à coefficients réels et de taille d par d, on note $\lambda_1(A), \lambda_2, \ldots, \lambda_d(A)$ les valeurs propres de A triées par ordre décroissant.

1. Montrer que :

$$|\lambda_1(M') - \lambda_1(M)| \leq \|P\|.$$

   (Indice : nous avons défini et prouvé certaines propriétés de la norme matricielle $\|.\|_2$ en cours.)

2. Soit $i \in \{2, 3, \ldots, d\}$. Notons $\mathcal{E}_i^d$ l'ensemble de tous les sous-espaces de dimension i de $\mathbf{R}^d$. Montrer que pour toute matrice symmétrique A à coefficients réels et de taille d par d, la

i-ième valeur propre de A est donnée par :

$$\lambda_i(A) = \min_{E \in \mathcal{E}_i^d} \max_{v \in E^\perp \cap \mathcal{S}^{d-1}} v^T A v,$$

où $E^\perp$ est l'ensemble des vecteurs de $\mathbf{R}^d$ orthogonaux à tous les éléments de E et $\mathcal{S}^{d-1}$ est l'ensemble des vecteurs de $\mathbf{R}^d$ de norme euclidienne 1.

3. Montrer que :

$$\max_{i \in \{1,2,\ldots,d\}} |\lambda_i(M') - \lambda_i(M)| \leq \|P\|_2.$$