CORRECTION DU TD 6 : MAXIMUM DE VRAISEMBLANCE.

COURS D'APPRENTISSAGE, ECOLE NORMALE SUPÉRIEURE, AUTOMNE 2015

Jean-Baptiste Alayrac
jean-baptiste.alayrac@inria.fr

1. EXERCICE : MAXIMUM DE VRAISEMBLANCES DANS DES MODÈLES SIMPLES

1) Soit un n -échantillon $(x_1 \dots x_n)$ réalisation i.i.d de variables aléatoires Gaussiennes de densité $p(x; \mu, \sigma^2)$ où μ est la moyenne et σ^2 la variance.

Le but est d'estimer les paramètres $(\hat{\mu}, \hat{\sigma}^2)$ les plus vraisemblables ayant conduits aux observations $(x_1, \dots x_n)$. On commence par écrire la fonction de vraisemblance du modèle qui est une fonction de μ et σ^2 :

$$L(x_1, \dots x_n; \mu, \sigma^2) = \prod_{i=1}^n p(x_i; \mu, \sigma^2).$$

Le principe du maximum de vraisemblance recherche :

$$(\hat{\mu}, \hat{\sigma}^2) \in \operatorname{argmax}_{\mu, \sigma} L(x_1 \dots x_n; \mu, \sigma^2).$$

Pour simplifier le calcul, on préfère trouver le maximum d'une transformation strictement croissante de la fonction de vraisemblance (fonction à valeurs strictement positives dans notre cas) comme la log-vraisemblance. Dans notre cas la log-vraisemblance s'écrit :

$$f(\mu, \sigma^2) = \log(L) = \sum_{i=1}^n -\frac{(x_i - \mu)^2}{2\sigma^2} - n \log(2\pi\sigma^2).$$

Les dérivées partielles annulées en μ et σ^2 conduisent aux estimateurs suivants :

$$\hat{\mu} = \frac{\sum x_i}{n}, \quad \hat{\sigma}^2 = \frac{\sum_i (x_i - \hat{\mu})^2}{n}.$$

Remarque : Ces équations sont naturelles au sens où on tombe sur la moyenne et la variance empirique des données.

2) Par hypothèse, si on conditionne par rapport à X , à savoir si l'on fixe $X = x$, alors $Y - \beta^T x$ suit une loi Gaussienne centrée (moyenne 0) de variance σ^2 . On en déduit donc que Y suit une loi Gaussienne de variance $\mu(x) = \beta^T x$ et σ^2 .

3) On suppose que l'on dispose d'un n -échantillon $(x_1, y_1) \dots (x_n, y_n) \in \mathbb{R}^p \times \mathbb{R}$. On applique la méthode du maximum de vraisemblance pour estimer la variance σ^2 et le paramètre de moyenne β . Il est important de noter qu'il s'agit ici d'un modèle conditionnel car on modélise la dépendance de Y en fonction de X . Pour β , remarquons d'abord que que la log vraisemblance s'écrit :

$$f(\beta, \sigma^2) = \log(L) = \sum_{i=1}^n -\frac{(y_i - \beta^T x_i)^2}{2\sigma^2} - n \log(2\pi\sigma^2).$$

La maximisation de cette log-vraisemblance en β revient exactement à résoudre un problème de régression linéaire de $Y = (y_1, \dots y_n)^T$ sur X la matrice de design des données $(x_1, \dots x_n)$. L'estimateur du maximum

de vraisemblance de β est donc $\hat{\beta} = (X^T X)^{-1} X^T Y$. Vu le calcul de la question 1, l'estimateur de maximum de vraisemblance pour la variance est :

$$\hat{\sigma}^2 = \frac{\sum_i (y - \hat{\beta} x_i)^2}{n}.$$

Remarque intéressante. La version régularisée de la régression peut aussi se retrouver dans la version probabiliste étudiée aussi. Pour cela il faut ajouter une distribution sur les paramètres β que l'on apprend. Cela revient à ajouter un a priori sur les paramètres. On peut par exemple supposer que β suit une loi gaussienne de moyenne nulle et de variance $\frac{1}{\lambda}$. On est alors emmené à résoudre le maximum a posteriori et on retombe normalement bien sur les équations régularisées que nous avons vu lors des TPs précédents.