

MIGS

Intitulé de l'enseignement : Modélisation probabiliste et Algorithmique

Année : M2

Algorithme EM et Gradient Stochastique

Vous avez collecté les durées de vie t_1, \dots, t_n (en années) de $n = 19830$ ampoules issues de deux lots (1 et 2) différents. Vous avez sauvegardé ces données dans le fichier **temps.txt** mais l'information du lot $l_1, \dots, l_n \in \{1, 2\}$ correspondant à chacune des durées de vie t_1, \dots, t_n a été perdue et vous aimeriez la retrouver.

On modélise la distribution des durées de vie des ampoules des lots 1 et 2 par des lois Gamma $\Gamma(\nu_1, \theta_1)$ et $\Gamma(\nu_2, \theta_2)$. On rappelle que la densité d'une loi gamma $\Gamma(\nu, \theta)$ est

$$f_{\nu, \theta}(t) = \frac{1}{\theta^\nu \Gamma(\nu)} t^{\nu-1} e^{-\frac{t}{\theta}} \mathbf{1}_{t>0}.$$

On notera p la proportion inconnue d'ampoules du lot numéro 1. Les paramètres inconnus du modèle seront décrit par $\Theta = (p, \mu_1, \nu_2, \theta_1, \theta_2)$. On notera T la variable aléatoire modélisant la durée de vie d'une ampoule de l'échantillon et L la variable aléatoire modélisant le numéro du lot dont elle est issue.

- Mettre en place le cadre théorique de l'algorithme EM. On précisera notamment la fonction $F_{\Theta'}(\Theta)$ à minimiser pour mettre à jour les paramètres du modèle.
- Implémenter la méthode du gradient stochastique permettant de minimiser $F_{\Theta'}$.
Pour l'initialisation des paramètres on pourra exprimer (ν, θ) en fonction de la moyenne et de la variance d'une loi $\Gamma(\nu, \theta)$.
Variante : implémenter l'algorithme du recuit simulé.
- Implémenter l'algorithme EM, le tester d'abord sur des données simulés puis donner une estimation de la proportion d'ampoules du lot 1 et des durées de vie moyennes des lots.
- Comparer la vitesse de convergence avec l'algorithme consistant à mettre à jour la moyenne et la variance avec les estimateurs classiques de ces quantités.