

Voila un résumé de l'article :

Soient N points $x_1, \dots, x_N$, où $x_i \in R^D$, chacun labelé par T annotateurs.

Le label de l'annotateur t pour le i -ème point est noté y_i^t . Mais les labels des différents annotateurs peuvent ne pas être corrects donc on note z_i le véritable label.

On note alors les matrices $X = [x_1^T, \dots, x_N^T] \in R^{N \times D}$ et $Y = [y_1^1, \dots, y_1^T, \dots, y_N^1, \dots, y_N^T] \in R^{N \times T}$.

Le but est de produire une estimation du véritable label $Z = [z_1, \dots, z_N]$. L'article traite seulement le cas de la classification binaire. Ainsi y_i^t and z_i peuvent seulement prendre les valeurs 0 ou 1, suivant qu'ils se trouvent dans le 1er ou 2nd groupe.

Le but est alors de maximiser $p(y_i^t, z_i | x_i)$, et pour cela l'algorithme EM va être utilisé.

L'article pose alors :

$$p(z_i | x_i) = (1 + \exp(-\alpha^T x_i - \beta))^{-1}, \quad \alpha \in R^D, \beta \in R$$

avec

$$p(y_i^{(t)} | x_i, z_i) = (1 - \eta_t(x))^{y_i^t - z_i} * \eta_t(x)^{1 - y_i^t - z_i}$$

$$\eta_t(x) = (1 + \exp(-w_t^T x_i - \gamma_t))^{-1}, \quad w \in M_{D \times T}, \gamma \in R^T$$

$$\tilde{p}(z_i) = \prod_t p(y_i^{(t)} | x_i, z_i) * p(z_i | x_i)$$

En effet, comme les valeurs de y_i^t peuvent seulement prendre des valeurs binaires (0 ou 1), l'article pose un modèle bernoulli pour $p(y_i^{(t)} | x_i, z_i)$. En ce qui concerne $p(z_i | x_i)$, les auteurs disent qu'ils ont choisi une régression logistique par facilité, mais qu'elle peut très bien suivre une autre distribution...

le Calcul de $p(y_i^t, z_i | x_i)$ avec l'algorithme EM donne alors pour les auteurs :

- 1ère étape : E-Step : calcul de $\tilde{p}(z_i) = p(z_i | x_i, y_i)$ Or avec la formule de Bayes et de la distribution a posteriori, j'obtiens :

$$p(z_i | x_i, y_i) = \frac{\prod_t p(y_i^t | z_i, x_i) p(z_i | x_i)}{\sum_i \prod_t p(y_i^t | z_i, x_i) p(z_i | x_i)}$$

(Je pense que jusque là il ne doit pas y avoir d'erreur...)

Ensuite pour la 2ème étape :

- M-Step : maximisation de la fonction

$$f_{opt} = \sum_i \sum_t E_{\tilde{p}(z_i)} [\log p(y_i^{(t)} | x_i, z_i) + \log p(z_i | x_i)]$$

par rapport aux paramètres $\alpha, \beta, \gamma_t, w_t$.

Or pour moi, f_{opt} est alors égale à :

$$f_{opt} = \sum_i \sum_t \tilde{p}(z_i) [\log p(y_i^{(t)} | x_i, z_i) + \log p(z_i | x_i)]$$

Et l'article utilise alors l'algorithme LBFGS pour calculer ce maximum...or l'utilisation de cet algorithme requiert le calcul des gradients par rapport aux différents paramètres. L'article donne

seulement le résultat obtenu pour le calcul du gradient par rapport à α , et je n'arrive pas à trouver comme lui ! L'article trouve comme solution :

$$\frac{\partial f_{opt}}{\partial \alpha} \alpha \sum_i \frac{\Delta \tilde{p} * x_i * \exp(-\alpha^T x_i - \beta)}{(1 + \exp(-\alpha^T x_i - \beta))^2}$$

where $\Delta \tilde{p} = \tilde{p}(z_i = 1) - \tilde{p}(z_i = 0)$.

Or, voici mes calculs à moi :

$\tilde{p}(z_i)$ a été estimé dans la 1ère étape de l'algorithme EM, donc est considéré maintenant comme une constante.

$$\frac{\partial f_{opt}}{\partial \alpha} = \sum_i \sum_t p(\tilde{z}_i) * \frac{\partial}{\partial \alpha} [\log p(y_i^{(t)} | x_i, z_i) + \log p(z_i | x_i)]$$

$p(y_i^{(t)} | x_i, z_i)$ ne dépend pas de α , donc on a :

$$\frac{\partial f_{opt}}{\partial \alpha} = \sum_i \sum_t \tilde{p}(z_i) \frac{\partial}{\partial \alpha} [\log p(z_i | x_i)]$$

et

$$\frac{\partial}{\partial \alpha} [\log p(z_i | x_i)] = \frac{x_i \exp(-\alpha^T x_i - \beta)}{(1 + \exp(-\alpha^T x_i - \beta))}$$

donc,

$$\frac{\partial f_{opt}}{\partial \alpha} = \sum_i \sum_t \tilde{p}(z_i) * \frac{x_i \exp(-\alpha^T x_i - \beta)}{(1 + \exp(-\alpha^T x_i - \beta))}$$

Comme $\tilde{p}(z_i)$ and $\frac{x_i \exp(-\alpha^T x_i - \beta)}{(1 + \exp(-\alpha^T x_i - \beta))}$ ne dépend pas de T, on obtain finalement :

$$\frac{\partial f_{opt}}{\partial \alpha} = T * \sum_i \tilde{p}(z_i) * \frac{x_i \exp(-\alpha^T x_i - \beta)}{(1 + \exp(-\alpha^T x_i - \beta))}$$

Ce qui n'est pas proportionel au résultat de l'article !!!

