HMWK1

October 21, 2018

MLE of π

Let $(z_i)_{i\in[\![1,n]\!]}$ *iid*

The log-likelighood is given by :

$$\ell(\pi) = \sum_{m=1}^{M} n_m \log \pi_m \,, \;\; n_m = \sum_{i=1}^{n} \mathbb{1}_{\{z_i=m\}}$$

$-\ell$ being convex and as $\exists\pi \in [0,1]^M / \pi^T 1_M = 1$, by Slater's constraints qualification we have strong duality and we can address its dual problem given by

$$\max_{\lambda} \min_{\pi} \mathcal{L}(\lambda, \pi) \,, \;\; \text{where } \mathcal{L}(\lambda,\pi) = -\ell(\pi) + \lambda(\pi^T 1_M - 1)$$

$\mathcal{L}$ being convex w.r.t to π we can minimize it through its gradient :

$$\forall m, \; \frac{\partial \mathcal{L}}{\partial \pi_m} = 0 \Rightarrow -\frac{n_m}{\pi_m} + \lambda = 0 \Rightarrow \pi_m = \frac{n_m}{\lambda}$$

Plus, $\pi^T 1_M = 1 \Rightarrow \sum_{m=1}^{M} \frac{n_m}{\lambda} = 1 \Rightarrow \lambda = \sum_{m=1}^{M} n_m = n$, hence :

$$\forall m, \; \hat{\pi}_m = \frac{n_m}{n}$$

MLE of Θ

Let $(x_i)_{i\in[\![1,n]\!]}$ and $(z_i)_{i\in[\![1,n]\!]}$ *iid* and $\Theta = [\theta_{mk}] \in [0,1]^{M\times K}$

Conditional probability allow us to write the log-likelihood as :

$$\ell(\Theta, \pi) = \sum_{m=1}^{M} n_m \log \pi_m + \sum_{k=1}^{K}\sum_{m=1}^{M} q_{mk} \log \theta_{mk} \,, \;\; n_m = \sum_{i=1}^{n} \mathbb{1}_{\{z_i=m\}} \,, \;\; q_{mk} = \sum_{i=1}^{n} \mathbb{1}_{\{z_i=m, x_i=k\}}$$

Samely,

$$\mathcal{L}(\lambda, \Theta, \pi) = -\ell(\Theta,\pi) + \begin{pmatrix}\pi^T \Theta 1_K - 1 & \pi^T 1_M - 1\end{pmatrix} \lambda$$

Derivating w.r.t to π we obtain the same estimtor as previously.

For Θ, the derivation goes like :

$$\forall m, k, \; \frac{\partial \mathcal{L}}{\partial \pi_m} = 0 \Rightarrow -\frac{q_{mk}}{\theta_{mk}} + \lambda_1 \pi_m = 0 \Rightarrow \theta_{mk} = \frac{q_{mk}}{\lambda_1 \pi_m}$$

Once again, the constraints gives us $\lambda_1 = n$, hence :

$$\forall m, k, \; \hat{\theta}_{mk} = \frac{q_{mk}}{n\hat{\pi}_m} = \frac{q_{mk}}{n_m}$$

MLE of π

Let $(z_i)_{i\in[\![1,n]\!]}$ iid, the log-likelighood is given by : $\ell(\pi) = \sum_{m=1}^{M} n_m \log \pi_m\,,\;\; n_m = \sum_{i=1}^{n} \mathbb{1}_{\{z_i=m\}}$

$-\ell$ being convex and as $\exists \pi \in [0,1]^M / \pi^T 1_M = 1$, by Slater's constraints qualification we have strong duality and we can address its dual problem given by : $\max_{\lambda} \min_{\pi} \mathcal{L}(\lambda, \pi)\,,\;$ where $\mathcal{L}(\lambda,\pi) = -\ell(\pi) + \lambda(\pi^T 1_M - 1)$

$\mathcal{L}$ being convex w.r.t to π we can minimize it through its gradient : $\forall m,\; \frac{\partial \mathcal{L}}{\partial \pi_m} = 0 \Rightarrow -\frac{n_m}{\pi_m} + \lambda = 0 \Rightarrow \pi_m = \frac{n_m}{\lambda}$

Plus, $\pi^T 1_M = 1 \Rightarrow \sum_{m=1}^{M} \frac{n_m}{\lambda} = 1 \Rightarrow \lambda = \sum_{m=1}^{M} n_m = n$, hence : $\forall m,\; \hat{\pi}_m = \frac{n_m}{n}$

MLE of Θ

Let $(x_i)_{i\in[\![1,n]\!]}$ and $(z_i)_{i\in[\![1,n]\!]}$ iid and $\Theta = [\theta_{mk}] \in [0,1]^{M\times K}$. Conditional probability allow us to write the log-likelihood as : $\ell(\Theta,\pi) = \sum_{m=1}^{M} n_m \log \pi_m + \sum_{k=1}^{K}\sum_{m=1}^{M} q_{mk} \log \theta_{mk}\,,\;\; n_m = \sum_{i=1}^{n} \mathbb{1}_{\{z_i=m\}}\,,\;\; q_{mk} = \sum_{i=1}^{n} \mathbb{1}_{\{z_i=m, x_i=k\}}$

Samely, $\mathcal{L}(\lambda,\Theta,\pi) = -\ell(\Theta,\pi) + \begin{pmatrix}\pi^T\Theta 1_K - 1 & \pi^T 1_M - 1\end{pmatrix}\lambda\,,\;\; \lambda \in \mathbb{R}_+^2$

Derivating w.r.t to π we obtain the same estimtor as previously.

For Θ, the derivation goes : $\forall m,k,\; \frac{\partial \mathcal{L}}{\partial \theta_{mk}} = 0 \Rightarrow -\frac{q_{mk}}{\theta_{mk}} + \lambda_1 \pi_m = 0 \Rightarrow \theta_{mk} = \frac{q_{mk}}{\lambda_1 \pi_m}$

Once again, the constraints gives us $\lambda_1 = n$, hence : $\forall m,k,\; \hat{\theta}_{mk} = \frac{q_{mk}}{n\hat{\pi}_m} = \frac{q_{mk}}{n_m}$

LDA formulas

$$Y \sim \mathcal{B}(\pi),\;\; X|\{Y=i\} \sim \mathcal{N}(\mu_i, \Sigma).$$

$$\hat{\pi} = \frac{1}{N}\sum_{i=1}^{n} y_i$$

$$\forall j \in \{0,1\},\;\; \hat{\mu}_j = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{y_i=j\}} x_i}{\sum_{i=1}^{n} \mathbb{1}_{\{y_i=j\}}}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{y_i=0\}}(x_i-\hat{\mu}_0)(x_i-\hat{\mu}_0)^T + \mathbb{1}_{\{y_i=1\}}(x_i-\hat{\mu}_1)(x_i-\hat{\mu}_1)^T$$

$$p(y=1|x) = \frac{1}{2} \Leftrightarrow (\mu_0-\mu_1)^T\Sigma^{-1}(x + \frac{\mu_0+\mu_1}{2}) = \log\left(\frac{\pi}{1-\pi}\right)$$

QDA formulas

$$Y \sim \mathcal{B}(\pi),\;\; X \,|\, \{Y = i\} \sim \mathcal{N}\left(\mu_i, \Sigma_i\right)$$

$$\hat{\pi} = \frac{1}{N}\sum_{i=1}^{n} y_i$$

$$\forall j \in \{0,1\},\;\; \hat{\mu}_j = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{y_i=j\}} x_i}{\sum_{i=1}^{n} \mathbb{1}_{\{y_i=j\}}}$$

$$\forall j \in \{0,1\},\;\; \hat{\Sigma}_j = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{y_i=j\}}(x_i-\hat{\mu}_j)(x_i-\hat{\mu}_j)^T}{\sum_{i=1}^{n} \mathbb{1}_{\{y_i=j\}}}$$

$$p(y=1|x) = \frac{1}{2} \Leftrightarrow \frac{1}{2}\log\left(\frac{\det\Sigma_1^{-1}}{\det\Sigma_0^{-1}}\right) + \frac{1}{2}\left[(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) - (x-\mu_0)^T\Sigma_0^{-1}(x-\mu_0)\right] = \log\left(\frac{\pi}{1-\pi}\right)$$