AHMED MAZARI

HOMEWORK 1

I) Representations: variantes du bayesian naif.

1) - a priori distribution are not uniform.
- $P(y) \sim$ Bernouli

$$h(x) \propto \underset{y}{\operatorname{argmax}} \int_{\theta} p(X|Y;\theta)\, p(\mathcal{D}|\theta)\, p(\theta)\, p(y;\alpha)\, d\theta$$

$$\int \beta_y^{x} (1-\beta_y)^{1-x} \prod_{k=0}^{n_k=1} \beta_k^{n_{k_1}} (1-\beta_k)^{n_{k_0}} \beta_k^{\gamma_{k-1}} (1-\beta_k)^{\delta_{k-1}} \frac{\Gamma(\gamma+\delta)}{\Gamma(\delta)\Gamma(\gamma)} \alpha^{y} (1-\alpha)^{1-y} d\beta$$

$$\frac{\Gamma(\gamma+\delta)}{\Gamma(\delta)\Gamma(\gamma)} \alpha^{y} (1-\alpha)^{1-y} \int \beta_y^{x+n_{y_1}+\gamma-1} (1-\beta_y)^{-x+n\gamma_0+\delta_\beta} \prod_{k\neq y} \beta_k^{n_{k_1}+\gamma_{\beta-1}} (1-\beta_k)^{\delta_\beta - n_k}$$

$$\frac{\Gamma(x+n_{y_1}+\gamma_\beta)\, \Gamma(n_{y_0}+\delta_\beta+1-x)}{\Gamma(n_{k_1}+\gamma_\beta+n_{k_0}+\delta_\beta+1)} \prod_{k=y} \frac{\Gamma(n_{k_1}+\gamma)\,\Gamma(n_{k_0}+\delta_\beta)}{\Gamma(n_{k_1}+n_{k_0}+\gamma+\delta)}$$

$$= \frac{\Gamma(\gamma+\delta)}{\Gamma(\delta)\Gamma(\gamma)} \alpha^{y} (1-\alpha)^{1-y} \left[\frac{(n_{k_1}+\gamma_\beta)^{x}\, \Gamma(n_{k_1}+\gamma_\beta)\, (n_{k_0}+\delta_\beta)^{1-x}\, \Gamma(n_{k_0}+\delta_\beta)}{\Gamma(n_{k_1}+\gamma_\beta+n_{k_0}+\delta_{\beta_0}+1)} \right.$$

Explication de la démarche :

loi à priori conjugué :

Si $\mathbb{P}(\theta)$ est choisi de façon à ce que $\mathbb{P}(\theta|\mathbb{C})p(\theta)$ et $p(\theta)$ aient la même forme alors $\mathbb{P}(\theta)$ loi à priori conjugué

$$\underset{y}{\operatorname{argmax}}\ \mathbb{P}(x^*, y|\mathbb{C}) = \underset{y}{\operatorname{argmax}} \int_\theta \mathbb{P}(x^*, y, \theta/\mathbb{C})\, d\theta$$

$$= \underset{y}{\operatorname{argmax}} \int_\theta \mathbb{P}(x, y|\theta)\, \mathbb{P}(\theta|\mathbb{C})\, d\theta$$

$$= \underset{y=1 \ldots 2}{\operatorname{argmax}} \int_\theta \mathbb{P}(x/y, \theta)\, p(y)\, p(\theta, \mathbb{C})$$

$$= \underset{y}{\operatorname{argmax}} \int_\theta \mathbb{P}(x|y, \theta)\, \mathbb{P}(\theta|\mathbb{C})\, d\theta$$

$\cancel{p(\hat{\theta}) = \underset{\beta \in B}{\operatorname{argmax}}\ p(\mathbb{C}|B)\, \varphi(B/y)}$

$\cancel{\hat{\theta} = \operatorname{arg}}$

2

1 Estimateur ML :

$$p(x, \theta) = \frac{l_x!}{x_1!\, x_2! \ldots x_{x_m}!}\, \beta_1^{x_1}\, \beta_2^{x_2} \ldots \beta_{x_m}^{x_m} \quad \text{et} \quad \sum_{w=}^{x_W} \beta_m = 1$$

$$P(\mathbb{C},\theta) = \prod_{d=1}^{n_d} l_{x_d}! \prod_{w_1}^{n_{wd}} \frac{\beta^{x_{wd}}}{x_{wd}!} = \prod_{d=1}^{n_d} l_{x_d}! \left(\prod_{w_1}^{n_{wd}} \frac{\beta_w^{x_{wd}}}{x_{wd}!} \right)$$

$$f(\theta) = \log P(\mathbb{C};\theta) = \sum_{d=1}^{n_d} \log(l_{x_d}!) + \sum_{w=1}^{w_d} x_{wd} \log(\beta_w) - \log(x_{wd}!)$$

$$\hat{f}(\theta) = f(\theta) - \lambda \left(1 - \sum_{w=1}^{n_w} \beta_w\right)$$

$$\frac{\partial \hat{f}}{\partial \beta_w} = \sum_d \left[\frac{x_{wd}}{\beta_w} - \lambda \right] = 0 \quad (1)$$

$$\frac{\partial \hat{f}}{\partial \lambda} = 1 - \sum \beta_\lambda = 0 \quad (2)$$

$$(1)\ (2) \Rightarrow \left\{ \sum_d x_{wd} = n_d \lambda \beta_w \iff n_d \lambda \underbrace{\sum_{w=1}^{n_w} \beta_w}_{=1} = \sum_{W=1}^{n_w} \sum_{d=1}^{n_d} x_{wd} = n_d \lambda \right.$$

On reinject $\beta_w = \dfrac{\sum_d x_{wd}}{n_d \sum_d l_{x_\lambda}}$

2. MAP

$$\overbrace{P(\theta \mid x)}^{\text{posterior}} = p(x \mid \theta)\, p(\theta) \propto p(x \mid \theta) \underbrace{l(\theta)}_{\text{prior}}$$

$$\theta = (\beta_1, \dots, \beta_w) \sim Dir(\alpha)$$

$$\sum_{k=1}^{n_w} \beta_k = 1, \quad \alpha_0 = \sum_{i=1}^{n_w} \alpha_i$$

$$P(\beta \mid \alpha) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{n_w} \Gamma(\alpha_i)} \beta_1^{\alpha_1 - 1} \cdots \beta^{\alpha_{n_w} - 1}$$

(4)

$$\log(P(\beta \mid \alpha)) = \log \Gamma(\alpha_0) - \sum_{i=1}^{n_w} \log(\Gamma(\alpha_i)) + \sum_{w=1}^{n_w} (\alpha_w - 1) \log \beta_w \quad (3)$$

$$(A) = \mathbb{P}(\beta/\alpha)\left[\prod_{w=1}^{n_w}\prod_{d=1}^{n_d}\mathbb{P}(x_{w_d}|\beta)\right]$$

$$P(\hat{\theta}) = \underset{\beta\in B}{\arg\max}\ P(C|\beta)\,\alpha(\beta|\alpha)$$

$$= \underset{\beta\in B}{\arg\max}\ (A)$$

$$\hat{\theta}\ \alpha\quad \underset{\beta\in B}{\arg\max}\sum_{d}\sum_{n_w}\log P(x_{w_d}|\beta) + \log P(B/\alpha)$$

3. loi prédictive

$$\mathcal{L}(\beta) = \sum_{d}\log\left(\ell_{x_d\lambda}\right) + \left[\sum_{w=1}^{n_w} x_{w_d}\log\beta_w - \log(x_{w_d}!)\right] - \log\Gamma(\alpha_0) - \sum^{n_w}\Gamma(\alpha_i) + \sum_{w=1}^{n_w}(\alpha_w - 1)\log\beta_w + \lambda\left(1-\sum_{w=1}^{n_w}\beta_w\right)$$

$$\frac{\partial \ell_\beta}{\partial\beta_w} = \sum_{d}\left(\frac{w_\lambda}{\beta_w} - \lambda\right) + \frac{\alpha_w - 1}{\beta_w} = 0 \iff$$

$$n_d\,\lambda\,\beta_w = \left(\alpha_w - 1 + \sum_d w_d\right)$$

$$\Rightarrow \left(\sum_{w=1}^{n_w}\beta_w\right) n_d\lambda = \left(\sum_{w=1}^{n_w}\alpha_w\right) - n_w + \sum_{d,w} x_{w_d}$$

$$\frac{\partial \hat{l}(\beta)}{\partial \theta} = \sum_d \left(1 - \sum_{w=1}^{n_w} \beta_w\right) = 0$$

$$= n_d \left(1 - \sum_{w=1}^{n_w} \beta_w\right) \Leftrightarrow \begin{cases} \lambda = \alpha_0 - n_w + \sum_d l_{x_d} \\ \text{Lagrange} \\ \beta_w = \dfrac{\alpha_w - 1 + \sum_d x_{wd}}{\alpha_0 - n_w + \sum_{d=1} l_{xd}} \end{cases}$$

II Manipulation des MG

$$p(x_1, x_2, x_3, x_4) = p(x_4)\, p(x_3)\, p(x_1 / x_4)\, p(x_2 / x_4, x_3)$$

$$X_1 \perp\!\!\!\perp X_2 \mid X_4 \Leftrightarrow p(X_1, X_2 / X_4) = p(X_1 / X_4)\, p(X_2 / X_4)$$

$$p(X_1 / X_2 X_4) = p(X_1 / X_4)$$

$$p(X_2 \mid X_1, X_4) = p(X_2 / X_4)$$

sachant que $\sum_{X_3} p(X_1, X_2, X_3, X_4) = p(X_1, X_2, X_4) = p(X_4) \cdot p(X_1 / X_4)\, p(X_2 / X_4)$

$$\Rightarrow \frac{p(X_1, X_2, X_4)}{p(X_4)} = p(X_1, X_2 / X_4) = p(X_1 / X_4)\, p(X_2 / X_4)$$

$$p(X_1, X_2, X_4) = p(X_1 / X_4)\, p(X_2, X_4)$$

$$\Rightarrow \frac{p(X_1, X_2, X_4)}{p(X_2, X_4)} = p(X_1 / X_2, X_4) = p(X_1 \mid X_4)$$

$P(X_1, X_2, X_4) = P(X_2 | X_4)\, P(X_1 | X_4)$

$\Rightarrow \dfrac{P(X_1, X_2, X_4)}{P(X_1, X_4)} = P(X_2 | X_1, X_4) = P(X_2 | X_4)$

d'où $X_1 \perp\!\!\!\perp X_2 \mid X_4$,

Calcule $P(E = v \mid D = n)$

$P(E|D) = \dfrac{P(E,D)}{P(D)} = \dfrac{P(D | E=v, C)\, P(E=v, C)}{P(D=n)}$

$= \dfrac{0,99 \times P(E=v, C=n) + 0,92\, P(E=v, C=o)}{P(D=n)}$

$= \dfrac{(0,99 \times 0,05 \times 0,505) + (0.92 \times 0.95 \times 0,495)}{0,13}$

$= 0,62$