CheatSheet

Friday, November 4, 2022 10:12 AM

EM :

Mixture gaussienne :

- Θ = {param de la loi en fonction des classes, apriori π_i }

observation / classe caché / loi des x^h en fct de la classe / a priori sur la rep des classes

- $P_\Theta(x_i^o, x_i^h) = P_\Theta(x_i^o \mid x_i^h)\, P_\Theta(x_i^h)$

$$\propto \begin{pmatrix} \mathcal{N}(\mu_1, \sigma_1)\, \pi_1 \\ \mathcal{N}(\mu_2, \sigma_2)\, \pi_2 \end{pmatrix} \quad \text{Pour 2 classe}$$

- Q_i^{t+1} = "Proba $\forall$ variables cachés sachant observ."
= "Proba $\forall$ classes sachant observation"

$$= P_{\Theta^t}(x_i^h \mid x_i^o) = \frac{P_{\Theta^t}(x_i^h, x_i^o)}{P_{\Theta^t}(x_i^o)}$$

$P(X = x_i)$
$X \sim \mathcal{N}(\mu_i, \sigma_i^2)$
= une densité

$$\propto P_{\Theta^t}(x_i^o \mid x_i^h)\, P_{\Theta^t}(x_i^h) \propto \text{"}\mathcal{N}(\mu_i, \sigma_i)\text{"}\, \pi_i$$

Puis normaliser !

- Problème de maximisation : (étape M)

On a $LL(X, \theta) = \log L(X^o, \Theta)$, $X = (x_1, \ldots, x_m)$

$$LL(X, \Theta) = \log P_\Theta(x_i^o) = \sum_{i=0}^{m} \log\Big(\sum_{x_i^h \in X^h} P_\Theta(x_i^o, x_i^h)\Big)$$

$$= \sum \log\Big(\sum_{x_i^h \in X^h} Q_i(x_i^h)\, \frac{P_\Theta(x_i^o, x_i^h)}{Q_i(x_i^h)}\Big)$$

$$= \sum \log \left(\sum_{x_i^h \in X^h} Q_i(x_i^h) \frac{P_\theta(x_i, x_i^h)}{Q_i(x_i^h)} \right)$$

$LL =$ Par Jensen $=$

= Somme pour tout point, et pour toutes classe inconnue de ...

constante

modèle $\Rightarrow$ inconnu

$$= \sum_{i=1}^{n} \sum_{h \in C} Q_i^{t+1}(x_i^h) \log \left(\frac{P_\theta(x_i, x_i^h)}{Q_i^{t+1}(x_i^h)} \right)$$

$\underset{\Theta}{\arg\max} \, LL = \arg\max$

osef pour argmax

$$= \underset{\Theta}{\arg\max} \sum_{i=1}^{n} \sum_{h \in C} Q_i^{t+1}(x_i^h) \log(P(x_i, x_i^h))$$

$$= \arg\max \sum \sum Q_i^{t+1}(x_i^h) \log\left(\pi_i \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2} \right)$$

constante osef argmax

$$= \arg\max \sum \sum Q_i^{t+1}(x_i^h) \left[\log(\pi_i) - \frac{1}{2}\log(\sigma_1^2) - \frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 \right]$$

Pour 2 classes.

⚠ $\pi_1 = 1 - \pi_2 \Rightarrow$ Moins de boulot !

- $$\frac{\partial LL(x_i^o, \theta)}{\partial \pi_1} = \frac{\sum_{i=0}^{n} Q_i^{t+1}(1)}{\pi_1} + \frac{\sum Q_i^{t+1}(2)}{1 - \pi_1} = 0$$

- $$\frac{\partial LL(x_i^o, \theta)}{\partial \mu_1} = \sum_{i=0}^{n} Q_i^{t+1}(1) \frac{x_i - \mu_1}{\sigma_1^2} = 0$$

- $$\frac{\partial LL(x_i^o, \theta)}{\cdot} = \sum Q_i^{t+1}(1) \left[-\frac{1}{\cdot} + \frac{(x_i - \mu_1)^2}{\cdot} \right]$$

- $\dfrac{\partial LL(x_i^o, \theta)}{\partial \sigma_1} = \sum Q_i^{t+1}(1) \left[-\dfrac{1}{\sigma_1} + \dfrac{(x_i - \mu_1)^2}{\sigma^3} \right]$

• EM normal

- Q_i^{t+1} pour les x_i sans données manquante $= 1$

$$Q_i^{t+1}(x_i^o) = 1$$

- $P(x_i^o, x_i^h)$ est donné directement par un tableau de valeur initiale

- $LLP_\theta(x^o) = \sum_{i=0}^{n} \sum_{x_i^h \in X^h} Q_i(x_i^h) \log \left(\dfrac{P_\theta(x_i^o, x_i^h)}{\underbrace{Q_i(x_i^h)}_{\text{osef argmax}}} \right)$

1) dériver $\forall \theta_i \in \Theta$, $\dfrac{\partial LL}{\partial \theta_i} = 0$

2) Système d'équation pour retrouver les nouveaux param $(AX = B)$

⚠ penser aux trucs du type $\sum \theta_i = 1 \Leftrightarrow \theta_i = 1 - \ldots$

• A Posteriorie

$$\underset{\theta}{\operatorname{argmax}}\, P(\theta \mid X) = \underset{\theta}{\operatorname{argmax}} \frac{P(X \mid \theta) P(\theta)}{\underset{\text{constante}}{P(X)}}.$$

$$= \underset{\theta}{\operatorname{argmax}}\, \underbrace{P(X \mid \theta)}_{\text{vraissemblance}} \underbrace{P(\theta)}_{\text{apriori}} = \operatorname{argmax} L(X, \theta)\, \Pi(\theta)$$

$\underline{\text{Log}} \Rightarrow \operatorname{argmax} LL(X, \theta) + \log(\Pi(\theta))$

Remplacer par les valeurs numerique

Remplacer par les valeurs numerique puis developper a fond, sommet polynome à la fin

$x|\theta$

$$P(x|\theta) = \frac{P(x,\theta)}{P(\theta)}$$

Σ

$$P(\theta)$$

$$P(x,\theta) = P(x|\theta)P(\theta)$$

Σ

$$P(x)$$

$$P(\theta|x) = \frac{P(x,\theta)}{P(x)}$$

- Max a posteriori = Max de vraisemblance $\Leftrightarrow$ a priori Uniforme

- independance

- Independance conditionnelle : $P(X,Y|Z) = P(X|Z)P(Y|Z)$
  Trouver le tableau $P(X,Y|Z) = \frac{P(X,Y,Z)}{P(Z)}$ puis par marginalisation les tableaux $P(X|Z)$ et $P(Y|Z)$

- Independance 2 à 2 : Tous indep entre eux
- Independance mutuelle : + intersection de tous = produit de tous

- Generalité proba

- Bayes : $$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- $$\sum_A P(A|B) = \frac{\sum_A P(A,B)}{P(B)} = 1 \Rightarrow$$ tableau de proba ⚠

- $\sum_A P(A|B) = \frac{\sum_A P(A,B)}{P(B)} = 1 \Rightarrow$ tableau de proba
- Proba totale $P(X,Y) = \sum_Z P(X,Y,Z)$
- Loi bernoulli : $P(B=x_i) = p^{x_i}(1-p)^{1-x_i}$
- Variance Empirique : $\sum_{i=1}^{n} (x_i - \bar{X})^2$
- Variance théorique : $Var(X) = E((X-E(X))^2)$
  $= E(X^2) - E(X)$
- Espérance : $E(X) = \sum x_i p_i = \int_{\mathbb{R}} x f(x)\, dx$
- $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \Leftrightarrow \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0,1)$
- $X \sim \mathcal{N}(\mu, \sigma^2)$, $\alpha X + \beta \sim \mathcal{N}(\alpha\mu + \beta, \alpha^2\sigma^2)$
- $\frac{1}{n}(X_1 + \dots + X_n) \sim \mathcal{N}\left(\frac{1}{n}(\mu_1 + \dots + \mu_n), \frac{1}{n^2}(\sigma_1^2 + \dots + \sigma_n^2)\right)$

- Test stat

- $D = \frac{(\text{effectif réel} - \text{effectif théorique})^2}{\text{effectif théorique}} \overset{H_0}{\sim} \chi^2(\text{nb classe} - 1)$
- $P(\{D > c\} | H_0) < \alpha$
- Puissance du test : $P(\mathcal{R} | H_1) = 1 - P(\bar{\mathcal{R}}_{H_1} | H_1)$
- TCL : $X_1, \dots, X_n$ iid, $m = E(X_1)$, $\sigma = \sqrt{Var(X)}$

  $$Z_n = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - m) \xrightarrow[n \to +\infty]{\mathcal{L}} Z \sim \mathcal{N}(0,1)$$

$Z_n = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - m) \xrightarrow[n\to+\infty]{\mathcal{L}} Z \sim \mathcal{N}(0,1)$

• Student : $X_1, \dots, X_n$ iid $\sim \mathcal{N}(m, \sigma^2)$ $\left.\begin{matrix} m \\ \sigma^2 \end{matrix}\right\}$ inconnu

$$\frac{\sqrt{n}}{\sqrt{V_n}}(\bar{X}_n - m) \sim \mathcal{T}(n-1)$$

• Markov

- Homogène : Proba ne change pas avec le temps
- État transcient : = Pas recurrent <=> $P(\exists n \in \mathbb{N}, X_n = e \mid X_0 = e) \neq 1$
- État absorbant : $P(X_1 = e \mid X_0 = e) = 1$
- Periode = k : Revenir dans l'état en k étape $Pgcd(1,2,3,\dots) = 1$
- Irreductible : ∀ état la proba d'y revenir est non-nul = Fortement connexe, pas d'état final/absorbant

[• Irreductible + discret = unique μ

• Irreductible + 1 état apériodique = Tous apériodique

• Irreductible + apériodique = ergodique

[• ergodique + Homogène $\xrightarrow{LFGN}$ proba stationnaire μ
  - μ = simulation puis frequence
  - $\mu A \to$ CVG
  - $\mu A = \mu$: thm du point fixe $\sum \mu_i = 1$
  - $\forall A^k \xrightarrow[k\to\infty]{} \mu$

• Def Markov chain
  - processus stochastique
  - processus en état discret
  - temps discret
  - Prop markov : (= consigne)
  - Homogène :

• Formule

$M^{-1} = \frac{1}{|M|} com(M)^T$ $\qquad M = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$

$|M|$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{|M|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$\begin{pmatrix} g & h & i \end{pmatrix}$$

$$\operatorname{cof}(M) = \begin{pmatrix} +\begin{vmatrix} e & f \\ h & i \end{vmatrix} & -\begin{vmatrix} d & f \\ g & i \end{vmatrix} & + \\ \cdots & - & - \end{pmatrix}$$