CheatSheet

Wednesday, January 4, 2023 3:27 PM

I HMM

$\Pi = P(s_1 = i)$

$A = P(s_t = j \mid s_{t-1} = i)$

$B = P(x_t \mid s_t = i)$

$B = \begin{pmatrix} x_1 & x_2 & \cdots \\ & & \end{pmatrix}$

apprentissage = Baum-Welch | Si étiquette : Comptage

Décodage : Viterby | Évaluation : Forward

1) • Forward : Calculer la proba de $X = (x_1, \dots, x_m) = x_1^m$

But : Trouver $P(x_1^T \mid \lambda) = \sum P(x_1^T, s_1^T \mid \lambda)$ ← Problème de combinatoire

• $\alpha_0(j) = \Pi_j B_j(x_1)$

• $\alpha_t(j) = P(x_1^t, s_t = j)$

$j \to$ chaque état

$= b_j(x_t) \sum_i^{\text{état}} a_{ij}\, \alpha_{t-1}(i)$

$A = \begin{pmatrix} 0{,}99 & 0{,}01 \\ 0{,}01 & 0{,}99 \end{pmatrix}, \quad B = \begin{pmatrix} 0{,}8 & 0{,}1 \\ 0{,}2 & 0{,}9 \end{pmatrix}$

t	x_t	$\alpha_t(A)$ $(j=1)$	$\alpha_t(B)$ $(j=2)$
1	1	$0{,}99 \times 0{,}8 = 0{,}792$	$0{,}01 \times 0{,}1 = 0{,}001$
2	0	$0{,}2(0{,}99 \times 0{,}792 + 0{,}01 \times 0{,}001) = 0{,}156818$	$0{,}9(0{,}01 \times 0{,}792 + 0{,}99 \times 0{,}001) = 8{,}019 \times 10^{-3}$
3	1	$0{,}8(0{,}99 \times 0{,}156818 + 0{,}01 \times 8{,}019 \times 10^{-3}) \approx 0{,}1237$	$0{,}1(0{,}01 \times 0{,}156818 + 0{,}99 \times 8{,}019 \times 10^{-3}) \approx 9{,}5 \times 10^{-4}$

$\Sigma \approx 0{,}1247$

2) • Viterby : Seq Observation ⟹ Seq État

Pb : $\hat{s}_1^T = \underset{s_1^T}{\arg\max} \prod_{t=1} P(s_t \mid s_{t-1}) P(x_t \mid s_t)$

$\delta_1(i) = \Pi_i B_i(x_1)$ | $\Psi_t(j) = \underset{i \in [1,N]}{\arg\max}\, \delta_{t-1} a_{ij}$

$\delta_t(i) = b_j(x_t)\left[\max_i \delta_{t-1} a_{ij}\right]$

$\begin{pmatrix} \delta(1) & \delta(1) \\ \delta(2) & \delta(2) \end{pmatrix} \cdot A = \begin{pmatrix} \circ & \cdot \\ \circ & \vdots \end{pmatrix} \begin{matrix} 1 \\ 2 \end{matrix}$ argmax

max max

x_t		max 1	max 2	$\delta_t(A)$	$\delta_t(B)$
0		$0{,}792$	$0{,}001$	x	x
		max	max		
1	A	$0{,}792 \times 0{,}99 = \boxed{0{,}7840}$	$0{,}792 \times 0{,}01 = \boxed{0{,}00792}$	$0{,}7840 \times 0{,}2 = 0{,}1568$	$0{,}00792 \times 0{,}9 = 7{,}128 \times 10^{-3}$
	B	$0{,}001 \times 0{,}01 = 1_{10}^{-5}$	$0{,}001 \times 0{,}99 = 0{,}00099$		
		A	B		
0	A	$0{,}1568 \times 0{,}99 = \boxed{0{,}1552}$	$0{,}1568 \times 0{,}01 = 1{,}568_{10}^{-3}$	$0{,}1552 \times 0{,}8 = 0{,}12416$ (max)	$7{,}0567_{10}^{-3} \times 0{,}1 = 7{,}05_{10}^{-4}$
	B	$7{,}128 \times 0{,}01 \times 10^{-3} = /$	$7{,}128 \times 0{,}99 \times 10^{-3} = \boxed{7{,}0567_{10}^{-3}}$		

II Regression

1) Regression normal

$Y = \alpha + \beta X + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \sigma)$

$Y_i \sim \mathcal{N}(\alpha + \beta x_i, \sigma)$

• Fonction de cout

$F(\alpha, \beta) = \sum (y_i - \alpha + \beta x_i)^2 \rightsquigarrow \underset{\alpha, \beta}{\arg\min}\, F(\alpha, \beta)$

$\nabla F = 0 \iff \begin{cases} \bar{y} = \alpha + \beta \bar{x} \\ \sum_{i=1}^m x_i (y_i - (\alpha + \beta x_i)) = 0 \end{cases}$

$\iff \begin{cases} b = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \dfrac{\operatorname{cov}(X, Y)}{V(X)} \\ a = \bar{y} - \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \bar{x} = \bar{y} - \dfrac{\operatorname{cov}(X,Y)}{V(X)} \bar{x} \end{cases}$

• Max Vraissemblance

$Y = f(X) \overset{\text{exemple}}{=} \alpha X^2 + \beta X + \gamma + \varepsilon$

$Y_i \sim \mathcal{N}(\alpha x_i^2 + \beta x_i + \gamma, \sigma)$

$P(y_i \mid x_i, \theta, \sigma) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - f(x))^2}$

$L = \prod_i \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i - f(x_i))^2}$

$LL = \sum -\log(\sqrt{2\pi}\sigma) - \dfrac{1}{2\sigma^2}(y_i - f(x_i))^2$

$\arg\max \ell = \arg\max -\sum (y_i - f(x_i))^2$

$\nabla = 0 \iff \begin{cases} \sum 2x_i^2 (y_i - \alpha x_i^2 - \beta x_i - \gamma) = 0 \\ \sum 2x_i (y_i - \alpha x_i^2 - \beta x_i - \gamma) = 0 \\ \sum 2 (y_i - \alpha x_i^2 - \beta x_i - \gamma) = 0 \end{cases}$

$\iff A \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = B$

$\iff A \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = B$

• R^2 & Variance

$\sigma_y = \frac{1}{m} \sum (y_i - \bar{y})^2 = \frac{1}{m} \sum (\hat{y}_i + e_i - \bar{y})^2$

⟹ Développement + $\operatorname{cov}(e, \hat{y}) = 0$

$= \frac{1}{m} \sum (\hat{y} - \bar{y})^2 + \frac{1}{m} \sum (e_i)^2 =$ Var expliquée + Var erreur residuelle

$R^2 = 1 - \dfrac{\text{Variance Residuel}}{\text{Variance Totale}} = 1 - \dfrac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$

$= 1 - \dfrac{\sum (e_i)^2}{\sigma_y}$

• Multidim

$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$

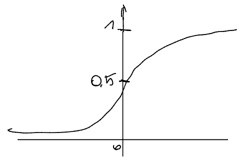
$\iff Y_i = [1 \ x_{i1} \ x_{i2}] \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \varepsilon_i = X\beta + \varepsilon$

$X = \begin{bmatrix} 1 & x_1^T \\ \vdots & \vdots \\ 1 & x_m^T \end{bmatrix}$ | $\hat{\beta} = (X^T X)^{-1} X^T Y$

$\sim \mathcal{N}(\beta, \sigma^2 (X^T X)^{-1})$

2) Regression Logistique

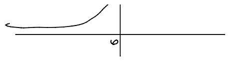
$P(Y = 1 \mid X = x) = f(x) = \dfrac{1}{1 + e^{-(x w + b)}}$



$L = \prod P(Y = y_i \mid X = x_i)$

$LL = \sum y_i \log(f(x_i)) + (1 - y_i) \log(1 - f(x_i))$

$\dfrac{\partial LL}{\partial w_j} = \sum x_{ij} (y_i - f(x_i))$



$\frac{\partial L}{\partial b} = \sum y_i - f(x_i)$

→ Montée de gradient OBLIGATOIRE

Algo Descente de gradient

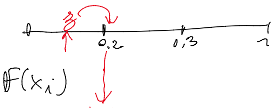
def grad (X, Y, max_iter=100, ε=, γ= $X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}$, $X_i \in \mathbb{R}^d$

```
N, d = X.shape              # Dimention
w = np.zeros((d,1))         # param
for _ in range(max_iter):
    w_0 = w.copy()
    w = w + ε x ∇f(X, w)
    if ((w - w_0)**2).sum() < gamma
        Break
Return w
```

II Sampling

1) Somme cumulée fct discrète

Cum sum des probas



$z \sim Unif([0,1[)$

$X_0 = 0$, $F(X_{i-1}) \leq z \leq F(X_i)$

Renvoyer X_i

On prend le x_i supp

2) Inversion de la fonction de rep

$F^{-1} :]0,1[\to \mathbb{R}$

$\alpha \to Inf\{t \in \mathbb{R}, F_x(t) \geq \alpha\}$

- Sur le graph de $F(x)$, placer α sur l'ordonnée
- Faire varier t sur chaque intervalle tq $F_x \geq \alpha$
- Puis prendre l'inf de la borne inférieure de cette intervalle

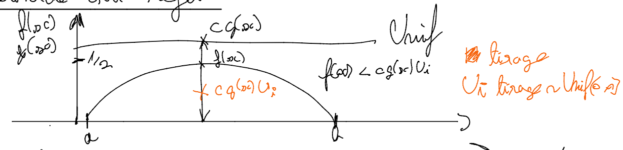
Exemple analytique : loi exp

$\forall \alpha \in]0,1[, F_x^{-1} = Inf\{t \in \mathbb{R}, F_x(t) \geq \alpha\}$

$= Inf\{t \in \mathbb{R}_+^*, 1 - e^{-\lambda t} \geq \alpha\}$

$= Inf\{t \in \mathbb{R}_+^*, t \geq \frac{\ln(1-\alpha)}{-\lambda}\} = -\frac{\ln(1-\alpha)}{\lambda}$

3) Méthode du rejet



$X \sim f(x)$, $g(x) \sim Unif[a,b]$

$= \frac{1}{b-a} 1_{]a,b[}(x)$? exemple pas fini

1) Tirer z selon $g(x)$
2) calculer $k\, g(z) = mq$
3) Tirer $u \sim Unif[0, mq]$
4) Accepter z ssi $u \leq f(z) = mq\, f(b(x))$

$P(\text{acceptés}) = \int \frac{g(x)}{k\, g(z)}$

4) MCMC

Construction d'une chaîne de Markov de loi stationnaire $\pi(\cdot)$

- Metropolis Hasting

$\pi(x) P(y|x) \alpha(x,y) = \pi P(x|y) \alpha(y,x)$

$\alpha(x,y) = \frac{\pi(y) P(x|y)}{\pi(x) P(y|x)}$ $\alpha(y,x) = 1$

Proba de realiser transition

- Generer x_{t+1} à partir de x_t
  1) tirer z selon $P(\cdot|x_t)$ (à choisir ?)
  2) $\alpha(x_t, z) = \min\{1, \alpha(x_t, z)\}$
  3) tirer $u \sim Unif[0,1]$
  4) $x_{t+1} = z$ (si $u \leq \alpha(x_t, z) / x_t$)

$P(z|x_t) = \begin{cases} q(\cdot) \to \text{random walk ?} \\ q(z) : \text{independant (n.s ?)} \end{cases}$

- Metropolis-Hastings par bloc

Generer $x_{t+1} = \{x_{t+1}^1 \dots x_{t+1}^n\}$ de x_t

1) Choisir σ : permutation $1, \dots, n$

$\forall i$ in σ :

a) $y = \{x_{t+1}^{\sigma(1)}, \dots, x_{t+1}^{\sigma(i-1)}, x_t^{\sigma(i+1)}, \dots, x_t^{\sigma(n)}\}$

b) tirer $z^{\sigma(i)} \sim P(\cdot | x_t^{\sigma(i)}, y)$

c) calculer $\alpha(x_t^{\sigma(i)}, z^{\sigma(i)} | y)$

d) Tirer $u \sim Unif[0,1]$

e) $x_{t+1}^{\sigma(i)} = \begin{cases} z^{\sigma(i)} \text{ si } u \leq \alpha \\ x_t \text{ sinon} \end{cases}$

4) Echantillonage de Gibbs

$P(z^{\sigma(i)} | x_t^{\sigma(i)}, y) = \pi(z^{\sigma(i)} | y)$

$x_{t+1}^{\sigma(i)} = z^{\sigma(i)}$