Compresión de Datos sin Pérdida

Procesos estocásticos

Álvaro Martín

[1]Instituto de Computación,
Facultad de Ingeniería
almartin@fing.edu.uy

[2]PEDECIBA Informática

Modelo de Markov de orden k

- $P(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots)$ solo depende de las últimas k observaciones: $x_{n-1}, x_{n-2}, \ldots, x_{n-k}$.
- Decimos que el símbolo x_n es *emitido* en el *estado* $s_n = x_{n-1}, x_{n-2}, \ldots, x_{n-k}$.
- *Conjunto de estados* $\mathcal{S} = \mathcal{X}^k$, de tamaño $|\mathcal{X}|^k$.
- Modelo definido una distribución de probabilidad para el *estado inicial* $s_1 = x_0, x_{-1}, \ldots, x_{-k+1}$, y una distribución de probabilidad sobre $\mathcal{X}$ condicionada en s, para cada $s \in \mathcal{S}$.
- En general, $P(s_1, x_1 \ldots x_n) = P(s_1) \prod_{i=1}^{n} P(x_i|s_i)$ y

$$P(x_1 \ldots x_n) = \sum_{s_1 \in \mathcal{S}} P(s_1) \prod_{i=1}^{n} P(x_i|s_i)\,.$$

- Para s_1 fijo, $P(x_1 \ldots x_n) = \prod_{i=1}^{n} P(x_i|s_i)$.

Relación con modelo de Markov de orden 1

- Podemos pensar un modelo de Markov de orden k sobre $\mathcal{X}$ como uno de orden 1 sobre $\mathcal{S} = \mathcal{X}^k$.
- *Matriz de prob. de transición* P de dimensiones $|\mathcal{S}| \times |\mathcal{S}|$.
- La emisión de a en estado $s_n = x_{n-1}, x_{n-2}, \ldots, x_{n-k}$ lleva al estado $s_{n+1} = a, x_{n-1}, \ldots, x_{n-k+1}$.
- Cada fila de P tiene no más de $|\mathcal{X}|$ entradas positivas.
- Para s_1 fijo, $x_1, x_2, \ldots, x_n$ determina $s_2, s_3, \ldots, s_{n+1}$ y viceversa. Por lo tanto, $H(X_1, X_2, \ldots, X_n) = H(S_2, S_3, \ldots, S_{n+1})$

Modelo FSM (máquina de estados finitos)

Modelo FSM para cadenas sobre $\mathcal{X}$ está definido por

- Un conjunto finito de estados, $\mathcal{S}$.
- Una distribución de probabilidad sobre $\mathcal{X}$ condicionada en s, $Q(\cdot|s)$, para cada $s \in \mathcal{S}$.
- Una distribución de probabilidad sobre $\mathcal{S}$, $Q(\cdot)$, para el *estado inicial* s_1.
- Una *función de próximo estado*, $f : \mathcal{S} \times \mathcal{X} \to \mathcal{S}$.

Una cadena de n símbolos se sortea siguiendo este procedimiento:

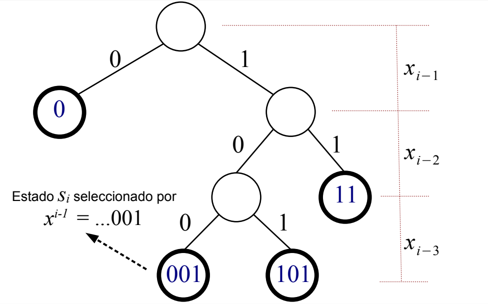
1. Se selecciona un estado inicial s_1 según $Q(\cdot)$.
2. Para cada i, desde $i = 1$ hasta n,
   2.1 Se sortea símbolo x_i según $Q(\cdot|s_i)$
   2.2 Se define $s_{i+1} = f(s_i, x_i)$.

En general, $Q(s_1, x_1 \ldots x_n) = Q(s_1) \prod_{i=1}^{n} Q(x_i|s_i)$.
Para s_1 fijo, $Q(x_1 \ldots x_n) = \prod_{i=1}^{n} Q(x_i|s_i)$.

Similitudes y diferencias

- Un modelo de Markov es un caso particular de Modelo FSM.
- Un modelo FSM define una cadena de Markov sobre el alfabeto de estados, $\mathcal{S}$.
- Su matriz de probabilidades de transición tiene dimensiones $|\mathcal{S}| \times |\mathcal{S}|$.
- $\mathrm{P}_{s,t} = Q(a|s)$, si existe $a \in \mathcal{X}$ t.q. $f(s, a) = t$, y $\mathrm{P}_{s,t} = 0$ en otro caso.
- Un modelo de Markov es *de memoria finita*: k símbolos del pasado determinan el estado actual de la cadena.
- Un modelo FSM no tiene por qué ser de memoria finita. La secuencia de estados es un proceso de Markov sobre $\mathcal{S}$, pero la secuencia de símbolos emitidos no necesariamente se pueden modelar como un proceso de Markov sobre $\mathcal{X}$.

Modelo árbol sobre $\mathcal{B} = \{0, 1\}$



- Modelo de memoria finita (Markov).
- Estados son hojas de un *árbol de contextos*.
- Potencialmente menos estados que Markov “plano”.
- No necesariamente determina una función de próximo estado.

Clase/Familia paramétrica de modelos

- Una *clase o familia* paramétrica de modelos sobre un alfabeto $\mathcal{X}$ es un conjunto
$$\mathcal{C} = \{P_\theta\}_{\theta \in \Theta}\,,$$
donde $\Theta \subset \mathbb{R}^d$ es un *espacio de parámetros* y $P_\theta : \mathcal{X}^* \to [0,1]$ es una función tal que, restringida a $\mathcal{X}^n$ es una distribución de probabilidad, para todo natural n y todo $\theta \in \Theta$. Es decir,
$$\sum_{x^n \in \mathcal{X}^n} P_\theta(x^n) = 1\,, \quad \theta \in \Theta.$$
- Ejemplos:
    - Familia de *Bernoulli*: i.i.d. sobre $\{0,1\}$, $\Theta = [0,1]$, $\theta = p(0)$.
    - Familia de modelos de Markov de orden k
    - Familia de modelos i.i.d. sobre $\mathcal{X}$
    - Familia de modelos determinados por una FSM fija.
    - Familia de modelos determinados por un árbol de contextos.

Tasa de entropía empírica

- Consideramos una clase paramétrica $\mathcal{C} = \{P_\theta\}_{\theta \in \Theta}$ y una secuencia $x^n \in \mathcal{X}^n$.
- Definimos la *máxima verosimilitud* de x^n como

$$P_{ML}(x^n) = \sup_{\theta \in \Theta} \{P_\theta(x^n)\}\,.$$

- Si $P_{\hat{\theta}}(x^n) = P_{ML}(x^n)$ le llamamos a $\hat{\theta}$ *parámetro de máxima verosimilitud* para x^n.
- La *tasa de entropía empírica* de x^n con respecto a $\mathcal{C}$ es

$$\hat{\mathcal{H}}(x^n) = -\frac{1}{n} \log P_{ML}(x^n)\,.$$

Ejemplo: Bernoulli

- Denotamos con n_0 y n_1 la cantidad de ocurrencias de 0 y 1, respectivamente en x^n.
- $P_\theta(x^n) = \theta^{n_0}(1-\theta)^{n_1}$.
- Tomando logaritmo natural obtenemos

  $$\ln P_\theta(x^n) = n_0 \ln \theta + n_1 \ln(1-\theta)\,,$$

  de donde derivando, e igualando a cero, despejamos $\hat{\theta} = n_0/n$.
- Entonces $P_{ML}(x^n) = (n_0/n)^{n_0}\,(n_1/n)^{n_1}$ y

  $$\hat{\mathcal{H}}(x^n) = -\frac{1}{n}\log P_{ML}(x^n) = -\frac{n_0}{n}\log\frac{n_0}{n} - \frac{n_1}{n}\log\frac{n_1}{n} = H(\hat{\theta})\,.$$
- Lo mismo ocurre para otras familias de modelos usuales.