

平成 30 年度知的情報処理システム特論 I レポート課題

次の問題 1~3 を解いて解答を提出せよ。

- 締切: 2018 年 7 月 6 日(金) 17:00
- 提出方法: 工学部 4 号館 604 号室の前においてある箱に提出する。レポートの一番上には、講義名、所属、学生証番号、名前を記入せよ。

問題 1

ある観測データに対し混合正規分布を仮定し、EM アルゴリズムによるパラメータ更新式を導出せよ。ただし、観測データ、混合正規分布、求めるべきパラメータは次のとおりとする。

- 観測データは $X = (x_1, x_2, \dots, x_N)$ とする。ただし、各 x_i は実数値(スカラー)とする。
- 混合正規分布は K 個の正規分布から成るとする。
- k 番目の正規分布 \mathcal{N}_k は次のように定義する。(各 μ_k, σ_k は実数値)

$$\mathcal{N}_k(x) = \mathcal{N}(x; \mu_k, \sigma_k^2) = \frac{1}{(2\pi\sigma_k^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma_k^2}(x - \mu_k)^2\right\}$$

- データ全体に対する混合正規分布

$p(y_k)$ を k 番目の正規分布が生成される確率とし、 $p(y_k) = \lambda_k$ とおく。

$p(x|y_k) = \mathcal{N}_k(x)$ であるため、データ全体の確率分布は、

$$p(X) = \prod_{i=1}^N \sum_{k=1}^K p(y_k) p(x_i|y_k) = \prod_{i=1}^N \sum_{k=1}^K \lambda_k \mathcal{N}_k(x_i) = \prod_{i=1}^N \sum_{k=1}^K \lambda_k \mathcal{N}(x_i; \mu_k, \sigma_k^2)$$

となる。ただし、 $0 \leq \lambda_k \leq 1$ (for all k) であり、 $\sum_{k=1}^K \lambda_k = 1$ である。

- 求めるべきパラメータ更新式: $\lambda_k, \mu_k, \sigma_k^2$

問題 2

HMM の解析問題について考える。状態集合 $Q = \{p, q, r\}$ 、記号集合 $\Sigma = \{0, 1\}$ としたとき、 π, a, b が次のように定義されているとする。

π		
p	q	r
0.46	0.10	0.44

a			
遷移元 \ 遷移先	p	q	r
p	0.55	0.18	0.27
q	0.13	0.05	0.82
r	0.54	0.09	0.37

b		
状態 \ 記号	0	1
p	0.78	0.22
q	0.68	0.32
r	0.59	0.41

このとき、入力 $x = "01100001101110111"$ に対する最も確率の高い状態列およびその状態記号列に対する確率を求めよ。

問題 3

ラベル集合を Q 、入力を x 、出力を $q_1 q_2 \dots q_T$ としたとき、Linear-chain CRF は次のように定義される。

$$p(q_1 q_2 \dots q_T | x) = \frac{1}{Z_x} \prod_{t=1}^T \exp \left\{ \sum_j \lambda_j f_j(q_{t-1}, q_t, x, t) \right\}$$

$$Z_x = \sum_{q'_1 \in Q, \dots, q'_T \in Q} \prod_{t=1}^T \exp \left\{ \sum_j \lambda_j f_j(q'_{t-1}, q'_t, x, t) \right\}$$

ただし、 $q_0 = DUMMY$ (ダミーのラベル) とし、素性関数 f の引数は次のように定義される。

f (時刻 $t-1$ のラベル, 時刻 t のラベル, 入力, 時刻)

また、 λ_j は f_j に対する重みとする。前向きスコア α と後ろ向きスコア β を次のように定義する。

$$\alpha(t, q) = \sum_{q_1 \in Q, \dots, q_{t-1} \in Q} \left(\prod_{u=1}^{t-1} e^{\sum_j \lambda_j f_j(q_{u-1}, q_u, x, u)} \right) e^{\sum_j \lambda_j f_j(q_{t-1}, q, x, t)}$$

$$\beta(t, q) = \sum_{q_{t+1} \in Q, \dots, q_T \in Q} e^{\sum_j \lambda_j f_j(q, q_{t+1}, x, t+1)} \left(\prod_{u=t+2}^T e^{\sum_j \lambda_j f_j(q_{u-1}, q_u, x, u)} \right)$$

このとき、ある時刻 t のラベル q_t に対する周辺確率 $p(q_t | x)$ を α, β を用いて求めよ。ただし、

$$p(q_t | x) = \sum_{q_1 \in Q, \dots, q_{t-1} \in Q} \sum_{q_{t+1} \in Q, \dots, q_T \in Q} p(q_1 \dots q_T | x)$$

である。