SOM のアルゴリズム

2019 年 8 月 5 日

表 1　変数記号表

記号	
$\mathbf{X}$	SOM のデータ集合（$\mathbf{X} = (x_{nd}) \in \mathbb{R}^{N \times D}$）
	$\boldsymbol{x}_n = (x_{n1}, \ldots, x_{nD})$
N	データ数
D	データの次元数
$\mathbf{Y}$	SOM の参照ベクトル集合（$\mathbf{Y} = (y_{kd}) \in \mathbb{R}^{K \times D}$）
	$\boldsymbol{y}_k = (y_{k1}, \ldots, k_{kD})$
$\mathbf{Z}$	潜在変数集合（$\mathbf{Z} = (z_{nl}) \in \mathbb{R}^{N \times L}$)
	$\boldsymbol{z}_n = (z_{n1}, \ldots, z_{nL})$
L	潜在空間の次元数
K	潜在空間のノード数
k_n^*	n 個目のデータ $\boldsymbol{x}_n$ の勝者ノード
ζ_k	k 番目のノードの潜在空間における座標
h_{kn}	K 個のノードと N 個の勝者ノードの組み合わせに対する学習率
T	総学習回数
τ	時定数 $(T > \tau)$
σ_{max}	近傍半径の最大値
σ_{min}	近傍半径の最小値
$\sigma(t)$	時刻 t における近傍半径 $\sigma(t) = max(\sigma_{max} - (\sigma_{max} - \sigma_{min})\frac{t}{\tau}, \sigma_{min})$
	$max()$ 演算子は入力集合の最大値を出力として返す.

1　SOM のアルゴリズム

1.1　人工データの作成

$\mathbf{X}, N, D$ の決定

1.2　ハイパーパラメータの設定

1.2.1　近傍半径のスケジューリングの設計

上記関数を用意.

1.2.2　離散化した潜在空間の設定

- ζ の範囲 ($[-1, 1]$ を推奨).
- ノード数 K の指定.
- 離散空間の次元数の決定

1.3　SOM の学習

以下を学習回数 T 回繰り返す.

■潜在変数の推定

$$k_n^* = \arg\min_k \|\boldsymbol{x}_n - \boldsymbol{y}_k\|^2 \tag{1}$$

$$\boldsymbol{z}_n := \boldsymbol{\zeta}_{k_n^*} \tag{2}$$

■参照ベクトルの推定

$$h_{kn} = \exp\left(-\frac{1}{2\sigma(t)^2}\|\boldsymbol{\zeta}_{k_n^*} - \boldsymbol{\zeta}_k\|^2\right) \tag{3}$$

$$g_k = \sum_n h_{kn} \tag{4}$$

$$\boldsymbol{y}_k = \frac{1}{g_k}\sum_n h_{kn}\boldsymbol{x}_n \tag{5}$$

1.4　ペアプログラミング

同級生 or 先輩 (放物双曲面: 宮崎, 曲線:瀬野浦) と潜在変数・参照ベクトルの数値を np.allclose で比較

1.5　描画

結果を matplotlib で描画. (funcanimation 推奨)