Apprentissage (III)

Mirta B. Gordon
Laboratoire Leibniz-IMAG
Grenoble

Systemes Complexes en SHS

ministère jeunesse éducation recherche — ministère délégué recherche et nouvelles technologies

CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

apprentissage non supervisé

clustering

- problème : classer (partager en sous-ensembles : clusters) un ensemble de données $\mathcal{L}_M = \{ \mathbf{x}^\mu \}_{1 \le \mu \le M}$
- critère : la similarité entre les données
  - il faut connaître le degré de similarité (distances) entre toutes les paires de points
- plusieurs algorithmes
  - K-means :
    - on se donne a priori le nombre de classes K
    - on cherche K hypersphères qui partagent au mieux les points
  - classifications hiérarchique :
    - on cherche le meilleur arbre (hiérarchie) en fonction des degrés de similarité (ne donne pas la classe des données)
  - super-paramagnetic clustering
    - la classe et le nombre de classes donnés par l'algorithme, à ≠ granularités

SPC (Blatt, Wiseman & Domany, 1996)

- soit $J_{\mu\nu}$ une fonction positive fortement décroissante de la distance entre les points $\mathbf{x}^\mu$ et $\mathbf{x}^\nu$
- on attribue à chaque point $\mathbf{x}^\mu$ une étiquette s_μ pouvant prendre un grand nombre de valeurs discrètes
- fonction de coût : $E = -\sum_{\substack{\mu,\nu=1 \\ (\mu \neq \nu)}}^{M} J_{\mu\nu}\, \delta_{s_\mu, s_\nu}$

delta de Kroenecker

$$\delta_{s_\mu, s_\nu} = 1 \text{ si } s_\mu = s_\nu$$

$$\delta_{s_\mu, s_\nu} = 0 \text{ sinon}$$

- modèle de Potts : s_μ = spin de Potts

propriétés

$$E = -\sum_{\substack{\mu,\nu=1 \\ (\mu\neq\nu)}}^{M} J_{\mu\nu}\,\delta_{s_\mu,s_\nu}$$

- tous les points interagissent, mais leur interaction décroit avec la distance
- si tous les s_μ sont égaux, E=minimum
- si $s_\mu \neq s_\nu$ la contribution $J_{\mu\nu}$ disparaît : l'énergie augmente
- au minimum : tous les points auront le même spin (la même étiquette)

clustering

$$E = -\sum_{\substack{\mu,\nu=1 \\ (\mu\neq\nu)}}^{M} J_{\mu\nu}\,\delta_{s_\mu,s_\nu}$$

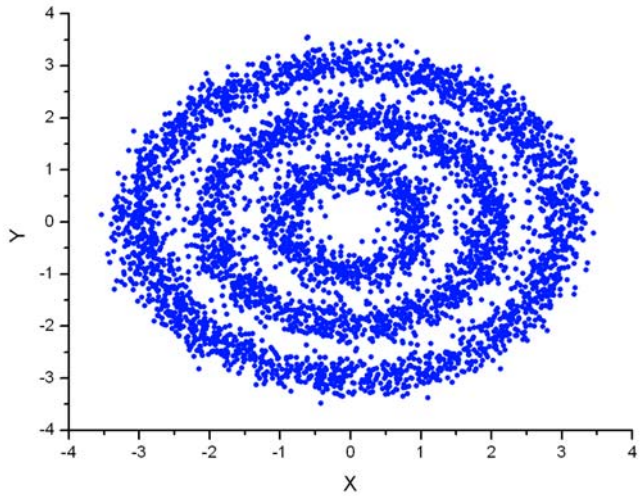
- évolution Montecarlo à temperature T=1/β :

$$\Delta E = E(s'_\mu) - E(s_\mu)$$

$$p(s_\mu \to s'_\mu) = \begin{cases} 1 & \text{si } \Delta E < 0 \\ \exp(-\beta\Delta E) & \text{si } \Delta E > 0 \end{cases}$$

- si $J_{\mu\nu}$ est grand, les spins μ et ν resteront égaux plus souvent que si $J_{\mu\nu}$ est petit

exemple



apprentissage_figures\3Conc.OPJ

références

- Hertz, Krogh et Palmer. Introduction to the Theory of Neural Computation. Addison Wesley, 1991.
- Tom Mitchell. Machine Learning. McGraw Hill, 1997.
- SPC : Blatt, Wiseman & Domany, Physical Review Letters 76 (N°18) 1996.
- SVM : V. Vapnik. Statistical Learning Theory, Wiley, 1998.

Merci !