

Aprendendo um perceptrão de Ising capaz de classificar dados naturais

Trabalho final original

J. Sacramento

INESC-ID Lisboa,
Instituto Superior Técnico,
Campus Taguspark, Porto Salvo, Portugal

Recuperação de Informação, DEAEIC, 2011

Tarefa

No nosso trabalho consideramos a tarefa usual de classificação (múltiplas classes)

- ▶ É nos fornecido um conjunto de treino da forma

$$\mathcal{S} = \{(\mathbf{x}^1, y^1), \dots, (\mathbf{x}^\mu, y^\mu), \dots, (\mathbf{x}^M, y^M)\},$$

que consideramos amostrado (i.i.d.) de uma distribuição desconhecida que gostaríamos de modelar

- ▶ Cada \mathbf{x}^μ é uma observação de treino
- ▶ $1 \leq y^\mu \leq C$ é a classe correspondente
- ▶ Classes mutuamente exclusivas

Arquitetura: camada de classificação

Propomo-nos a resolver o problema utilizando uma arquitectura neuronal artificial

- ▶ A tarefa propriamente dita é resolvida numa camada de classificação caracterizada por uma matriz de pesos \mathbf{J} de dimensão $C \times m$
 - ▶ Arquitectura tradicional simples
 - ▶ Unidades capazes de aprender separações lineares
 - ▶ Um perceptrão para cada classe, cada unidade dá uma resposta bipolar $\{-1, 1\}$
 - ▶ Resposta final obtida por competição: unidade com campo $h_i = \sum_{j=1}^m J_{ij}x_j$ mais forte responde $+1$, todos os outros respondem -1

Aprendizagem

Existem vários algoritmos eficientes para a construção de perceptrões dado um conjunto de treino \mathcal{S} quando não colocamos restrições na matriz \mathbf{J}

- ▶ Resolução eficiente (i.e., em tempo polinomial em m) de problemas linearmente separáveis

Grande questão:

É possível resolver a tarefa com uma matriz binária?
(perceptrão de Ising)

Motivação

Motivação forte para a questão anterior

- ▶ Implementação robusta em paralelo, em circuitos simples
- ▶ Provavelmente mais próximo da biologia

Infelizmente, estamos perante um problema altamente não-trivial de natureza combinatória

- ▶ Formulação geral permite concluir que o problema é NP-difícil (Senn and Fusi, 2005; Huang and Zhou, 2011)
- ▶ Mecânica estatística extrapola existência do perceptrão de Ising (Krauth and Mézard, 1989)
- ▶ Arrefecimento revela superfície de erro muito rugosa (Horner, 1992)

Abordagem naive

Dada a disponibilidade de algoritmos eficientes para o perceprão real, uma primeira solução de boa fé consiste em tomar os sinais de uma matriz previamente aprendida

- ▶ Não permite aprendizagem online
- ▶ Pode não resultar

Intuições mecânicas levam à conclusão de que a estratégia ingénua pode resultar em determinadas situações teóricas (Penney and Sherrington, 1993; Reimers et al., 1996).

Proposta de método

- ▶ Camada adicional de pré-processamento introduzida
 - ▶ Dados esféricos são projectados com projecção linear aleatória, especificada por uma matriz binária $\mathbf{R} \in \{-1, 1\}^{d \times m}$
 - ▶ Resultado da transformação é passado através de uma não-linearidade $\text{sign}(\cdot)$ de forma a obter vectores no hiper-cubo $\{-1, 1\}^d$
- ▶ Utilização de regra offline (regressão linear OLS) bem conhecida para aprender \mathbf{J} a partir dos vectores de treino projectados
- ▶ Versão binarizada $\text{sign}(\mathbf{J})$ é usada na fase de classificação

N.B.

- ▶ Arquitectura resultante apenas com actividade bipolar, embora vectores de entrada pertençam a \mathbb{R}^m
- ▶ Parâmetro d controla a complexidade do classificador

Resultados

Testámos o método num conjunto de dados naturais bem conhecido (Reuters 21578)

- ▶ 'Sabor' ModApté, 10 primeiras classes
- ▶ Dados esferizados, normalização ℓ_2
- ▶ Comparação com método tradicional de procura do vizinho mais próximo

Classificador	# errados / # total
Rede proposta ($d = 0.5m$)	0.0835
Rede proposta ($d = 0.75m$)	0.0514
Rede proposta ($d = m$)	0.0449
Vizinho mais próximo	0.0831

Discussão

- ▶ Método obtido muito simples com fácil implementação de classificador paralelo
- ▶ Pré-processamento altamente não-linear induz distribuição artificial que favorece construção do perceptrão de Ising através de simples binarização
- ▶ Algoritmo offline pode ser substituído por componente mais poderosa (e.g., adatron, minover)
- ▶ Propriedades de generalização muito interessantes observadas num conjunto de dados reais

Agradecimentos

Gostava de agradecer ao meu amigo e colega Ângelo Cardoso (INESC-ID) pelos inúmeros comentários e por toda a ajuda prestada

References I

- H. Horner. Dynamics of learning for the binary perceptron problem. *Zeitschrift für Physik B Condensed Matter*, 86(2):291–308, 1992.
- H. Huang and H. Zhou. Combined local search strategy for learning in networks of binary synapses. *EPL (Europhysics Letters)*, 96(5):58003, 2011.
- W. Krauth and M. Mézard. Storage capacity of memory networks with binary couplings. *J. Phys. France*, 50(20):3057–3066, 1989.
- R. W. Penney and D. Sherrington. The weight-space of the binary perceptron. *Journal of Physics A: Mathematical and General*, 26(22):6173, 1993.
- L. Reimers, M. Bouten, and B. V. Rompaey. Learning strategy for the binary perceptron. *Journal of Physics A: Mathematical and General*, 29(19):6247, 1996.
- W. Senn and S. Fusi. Convergence of stochastic learning in perceptrons with binary synapses. *Physical Review E*, 71(6):061907, 2005.