Case Study: Predicting Breast Cancer Invasion with Artificial Neural Networks on the Basis of Mammographic Features

MEDINFO 2004,
T02: Machine Learning Methods for Decision Support and Discovery

Constantin F. Aliferis & Ioannis Tsamardinos
Discovery Systems Laboratory
Department of Biomedical Informatics
Vanderbilt University

Motivation

- Lo, Baker, Kornguth, Iglehart, Floyd, Yearbook of Medical Informatics 1998
- Predict breast cancer invasion in order to prevent biopsies of benign lesions
- Decrease the cost and mortality of detecting small breast cancers in a screening population
- In a previous study the authors developed a NN that differentiates between benign vs cancerous lesions
- In this study, their aim was to develop a NN to differentiate, among the cancerous, between invasive and in situ

Data Collection

- Patients: 254 randomly collected women out of those who underwent needle localization of nonpalpable breast leasions
- 266 lesions sampled at open excisional biopsy to attain a definitive histopathologic diagnosis
- 170 benign lesions
- 96 cancers
  - 68 invasive
  - 28 in situ
- Only the 96 cancers were used to differentiate between invasive and in situ (to obviate diagnostic excisional biopsy of invasive cancers)

Observed Mammographic Quantities

- Calcification Distribution: discrete 0-5
- Calcification Number: discrete 0-3
- Calcification Description: discrete 0-14
  - E.g., vascular $\rightarrow$ 4
  - Spherical $\rightarrow$ 5
- Mass Margin: discrete 0-5
- Mass Size: real mm
- Mass Shape: discrete 0-4
- Mass Density: discrete 0-4
- Associated Findings: 0-9
  - E.g., hematoma $\rightarrow$ 2
- Special Cases
  - E.g., asym. breast tissue $\rightarrow$ 2
- Age: integer years
- 10 variables all-together
- Variables (except mass size and age) normalized between 0 and 1 (overfitting?)

Artificial Neural Network

- One hidden layer
- One output unit
- 15 hidden units
- Training algorithm: back-propagation with momentum 0.2 and learning rate 0.3 and 0.2 for the hidden and output layers respectively (why different?)
- “Network parameters [number of layers, units, learning rate] were optimized empirically”!!!!
- “The network performance was maximized after training for 4,000 iterations [epochs]” (no details on stopping criterion, no weight decay, or how 4000 was chosen)
- Leave-one-out evaluation of performance
- Transfer function? (sigmoid?)
- Training time ~4 hours (no details on the computer). On modern software and hardware should take less than 10secs

Performance

- Performance measure Area Under the ROC curve
- Area under ROC = .91 $\pm$ .03 (estimated by Leave One Out)
- Network built on all cases=Area under ROC .997 $\pm$ 0.004 (is this indeed the expected AUC?)

Conclusions

- A predictive model built to differentiate in women with breast cancer, the invasive from in situ
- Neural networks were used and trained with standard algorithms and parameters
- Unknown how many different parameters and encodings were used
- Performance of the final model very encouraging
- Model may help in reducing cost and danger from biopsies

Discussion

- Overfitting?
- Underfitting?
- Other encodings?
- Multi-classification?
- Decision analysis using cost of misclassification
- Other?