

Artificial Neural Networks and Ranking Approach for Probe Selection and Classification of Microarray Data

CMPE 59H - BIOINFORMATICS
BUSE BUZ
ONUR POYRAZ

Introduction

- ▶ We already know that there is a strong relationship between gene structure and some cancer types.
- ▶ In this study gene expression values come from cancer patients with either acute lymphocytic leukemia (ALL) or acute myeloid leukemia (AML).
- ▶ In the study, there is a dataset which used to classify which type of cancer each patient had based on measurements of their gene expressions.
- ▶ The study was published in 2013 and try to improve Golub's work (1999) which was the first to show that cancer types can be determined based on gene expressions alone.

Dataset

- ▶ Each row represents a different gene.
- ▶ First two columns are descriptions about that gene and each numbered column is a patient.
- ▶ Each patient has 7129 gene expression values - i.e each patient has one value for each gene
- ▶ The **training** data contain gene expression values for patients 1 through 38
- ▶ The **test** data contain gene expression values for patients 39 through 72
- ▶ They only used 50 gene expressions in order to find the classification

Dataset

	Gene Description	Gene Accession Number	1	call	2	call.1	3	call.2	4	call.3	...	29	call.33
0	AFFX-BioB-5_at (endogenous control)	AFFX-BioB-5_at	-214	A	-139	A	-76	A	-135	A	...	15	A
1	AFFX-BioB-M_at (endogenous control)	AFFX-BioB-M_at	-153	A	-73	A	-49	A	-114	A	...	-114	A
2	AFFX-BioB-3_at (endogenous control)	AFFX-BioB-3_at	-58	A	-1	A	-307	A	265	A	...	2	A
3	AFFX-BioC-5_at (endogenous control)	AFFX-BioC-5_at	88	A	283	A	309	A	12	A	...	193	A
4	AFFX-BioC-3_at (endogenous control)	AFFX-BioC-3_at	-295	A	-264	A	-376	A	-419	A	...	-51	A

Methods

- ▶ Feature selection for finding more related genes
 - ▶ GINI : Measures the inequality among features of a frequency distribution –best for in our case
 - ▶ Rank features
 - ▶ Fisher scoring
 - ▶ Kruskall
 - ▶ Stepwise regression :
- ▶ Artificial Neural Network for classification

Methods

- ▶ We are going to use these models and compare all of them,
 - ▶ PCA to reduce the dimension of feature set
 - ▶ K means clustering to select best match features
 - ▶ Bayesian optimization
 - ▶ Multilayer Perceptron
 - ▶ SVMs

References

- ▶ Artificial Neural Networks and Ranking Approach for Probe Selection and Classification of Microarray Data
AW Faria et al.
2013 BRICS Congress on Computational Intelligence & 11th Brazilian Congress on Computational Intelligence
- ▶ Monitoring Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression
T. R. Golub et al.
Science DOI: 10.1126/science.286.5439.531