Library of Models:
A New Methodology for Temporal Predictive Modeling in Biomedicine
Application in Predicting Laboratory Test Abnormalities

Ioannis Tsamardinos

Department of Biomedical Informatics, Vanderbilt Univ.

Collaborators:

Constantin F. Aliferis

S. Trent Rosenblum

Alexander Statnikov

In a Nutshell

- Developed a new modeling methodology to deal with heterogeneous populations of patients (in terms of available information about them) and learn predictive models; Library of Models
- Applied it to a real clinical problem (preliminary but extensive study): the prediction of laboratory abnormalities
  - 16848 models entered the library (84240 models were created in total to optimize parameters); a year's single-CPU time
  - Evaluated their performance prospectively to a year's worth of laboratory tests
- Promising initial results
- Promising future directions

Laboratory Tests

- Necessary for optimal diagnosis and treatment selection
- The use of laboratory tests increased in the last decades
- But…
- “Inappropriate” testing also increased
- Inappropriate testing a primary reason for the overall increase of testing
- 10-50% of tests classified as inappropriate in most studies

Why do Doctors Perform Inappropriate Tests

- "Unfortunately, physicians order laboratory tests for many actual reasons in addition to the usual accepted purposes of diagnosis, monitoring, screening, prognosis, and confirmation of clinical opinion. These include pressure from patient, family, or peers; hospital or legal requirements; defensive or medical legal protection; curiosity, insecurity, or delaying tactics; reassurance of themselves, the patient, the family; profit for the hospital, managed care company, laboratory, or themselves; to establish a baseline, complete a database, or having frustration at nothing else to do; ease of performance with ready availability because somebody will pay for it; or, probably most frequent of all, habit"

[The need for an outcomes research agenda for clinical laboratory testing, Lundberg, JAMA, 280(6), 1998]

Problems of Inappropriate Testing

- Causes unnecessary risk and discomfort to the patient
- Increases cost of health care
  - Each test cheap, but their absolute number (and cost) is high
  - Delaying discharging an inpatient until a test is complete, incurs a high indirect cost
- Increases the chances of false positives
  - May lead to further increases in cost
  - May cause anxiety

Which Tests are Inappropriate

- Definitions in the literature
- A test is inappropriate if it does not change the treatment course
  - Test could have changed treatment course
  - A single test may not single-handedly affect treatment course
- (among other criteria) A test is inappropriate if results are normal
  - Only way to appropriate testing is to know a priori which results will turn out abnormal
  - If you can predict abnormalities, testing is redundant and inappropriate

Which Tests are Inappropriate

- Other definitions appeared that are test and context specific
  - E.g., a lab test is inappropriate if its value could not have changed since last time
- Some definitions are implicit; they require human (subjective) judgment and manual reviews of patient records
- Difficult to define inappropriate testing in a general and explicit way to automate checking

Interventions to Reduce Number of Tests

- No such a general definition
- Devise interventions and Decision Support that reduce the number of tests
- Presumably, reducing the number of tests, only reduces the number of inappropriate tests performed

Interventions to Reduce Number of Tests

- Dissemination of guidelines
- Lectures
- Distribution of educational material
- Utilization audits
- Presentation of previous laboratory results or laboratory charges
- Laboratory requisition form modification

Results of Current Intervention Studies

- Number of tests reduced
  - Depending on the intervention short-term effects or long-term effects were detected
- Unknown:
  - Were only inappropriate tests avoided?
  - Or, do the physicians yield under the pressure to reduce laboratory charges, or to conform to the average laboratory use in the institution and are not ordering laboratory tests that are indeed needed?

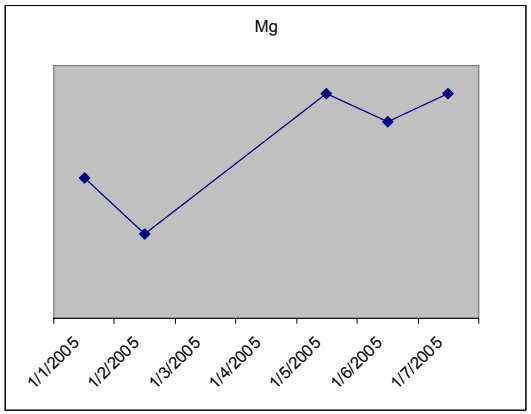
Our Approach to Inappropriate Testing

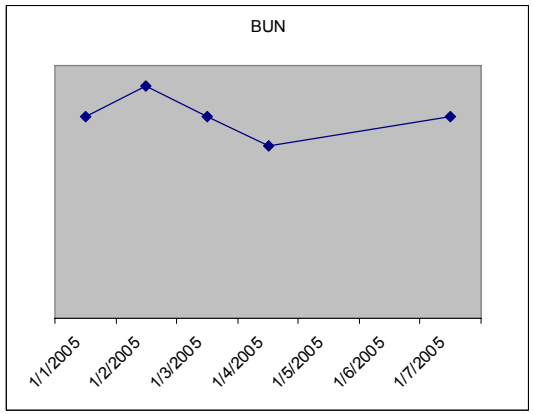
- Definition: A test is inappropriate if its outcome can be reliably predicted (lower bound on the number of inappropriate tests)
- Operational Application of Definition: Produce predictive models for laboratory test results
- Intervention: Design an intervention around the models (e.g., indicate the models' prediction before ordering)
- Evaluation: Measure the effects of the intervention, the reduction in test ordering, the accuracy of the prediction and quantify the benefits/losses monetarily and in terms of other health care outcomes

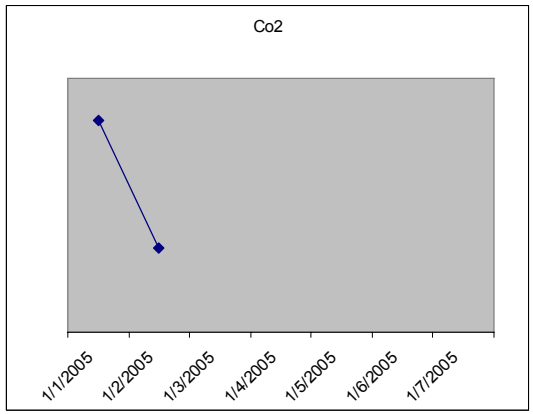
Predicting Laboratory Test Abnormalities

- Available data on a patient
  - Demographic
  - Prior hospitalization data
  - Prior laboratory results
- Predict whether the next laboratory test is going to be normal or abnormal

Patient Data







- Age: 33
- Gender: Male
- Admission Month: June
- Hospital Unit: cardiology
- Primary Diagnosis of Last Prior Hospitalization: 123.4 (ICD-9 code)
- Time since prior hospitalization: 1 year

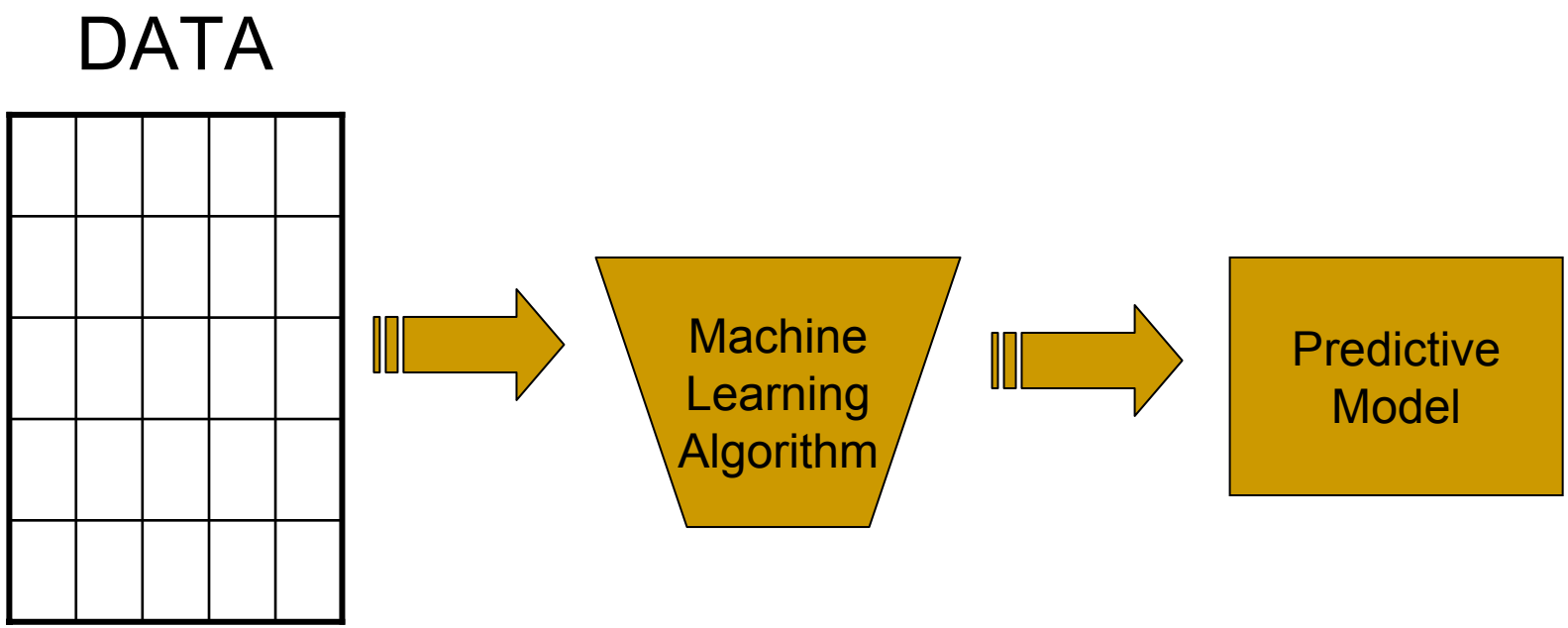
Question: Is the Mg value abnormal on the eighth day?

Standard Machine Learning and Statistical Modeling

	Age	Gender						Predictor *N*	Mg result abnormal ?
Mg test 1	33	Male							True
Mg test 2	24	Female							False
…									
Mg test *M*	3	Male							True

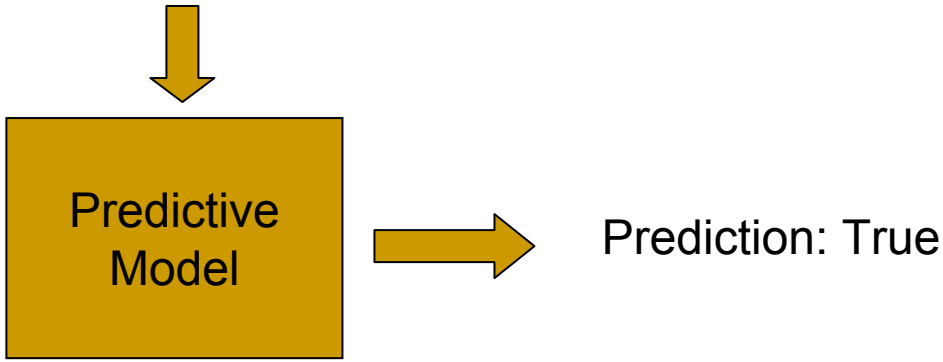
Fixed number and type of predictors is required

Standard Machine Learning and Statistical Modeling



Standard Machine Learning and Statistical Modeling

	Age	Gender	…	…	…	…	…	Predictor *N*	Next Mg result abnormal ?
New Lab Test	40	Male	…	…	…	…	…	…	??



Modeling Temporal Information

- How can temporal and atemporal information be encoded for use by Machine Learning methods?
- Common situation in predictive modeling in biomedicine:
  - Temporal and atemporal predictors
  - Temporal information not sampled at regular intervals
  - Temporal information not sampled at the same time across temporal predictors and patients

Different Encodings in Comparison

- Stationary vs. Non-stationary
  - Non-Stationary builds a model for a specific day (after hospitalization)
    - Predict Mg on Day 5
    - Stationarity assumes absolute time does not matter
- Absolute vs Relative Time Encoding
  - Absolute time encoding pre-specifies a number of days, it then uses the lab results on these days as predictors
  - Relative time encoding pre-specifies a number of previous measurements (whether yesterday or a month ago). It then encodes the results of the *i*th previous measurement and how long ago it was taken

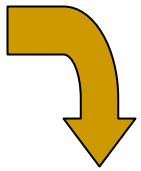
Stationary – Absolute Time Encoding

- Define a time interval (e.g. a day)
- To create a training sample for predicting the next lab using the measurements of *P* previous days
- For each temporal variable (predictor) *X (e.g.* Magnesium lab value)
  - define an atemporal variable X_i *(i=1,…,P-1)* denoting the value of *X i* days ago (before prediction is given)
- If more than one values are available for a given day *i*, X_i is the mean (or max, or min) of all the values of day *i*

Stationary – Absolute Time Encoding

A patient case for learning to predict the next Magnesium lab given the results of 2 previous days for each lab

Patient X	
Mg, Day 1	1.7
Mg, Day 2	Not Performed
Ca, Day 1	Not Performed
Ca, Day 2	9.0
Mg, Day 3	**Normal**



Matrix Encoding

	Mg_2	Mg_1	Ca_2	Ca_1	Mg
Patient X	1.7	Missing	Missing	9.0	**Normal**

All patients with a measurement after day 3 for Mg can be used as examples to train a predictive model for Mg using the measurements of the previous 2 days. A patient with a measurement on the fourth day can be used twice as a training example.

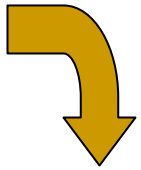
Stationary – Relative Time Encoding

- To create a training sample for predicting the next lab using the *P* previous measurements
- For each temporal variable (predictor) *X (e.g.* Magnesium lab value)
  - define an atemporal variable X_i denoting the value of *X, i* measurements ago
  - define an atemporal variable TX_i denoting how long ago measurement X_i was taken

Stationary – Relative Time Encoding

A patient case for learning to predict the Magnesium given 1 previous measurements of each lab

Patient X	
Mg, Day 1	1.7
Mg, Day 2	Not Performed
Ca, Day 1	Not Performed
Ca, Day 2	9.0
Mg, Day 3	**Normal**



Matrix Encoding

	Mg_1	TMg_1	Ca_1	TCa_1	Mg_3
Patient X	1.7	2 days	9.0	1 day	**Normal**

All patients with at least 2 measurements for Mg can be used as examples to train a predictive model for Mg using one prior measurement for all labs. A patient with 3 (and more) measurements can be used twice (and more) as a training example

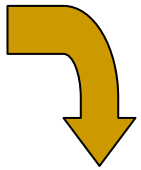
Non Stationary – Absolute Time Encoding

- Define a time interval (e.g. a day)
- To create a training sample for predicting a lab on Day *K* using the measurements of the previous *K-1 days*
- For each temporal variable (predictor) *X (e.g.* Magnesium lab value)
  - define atemporal variables X_i *(i=1,…,K-1)* denoting the value of *X* on day *i* from hospitalization
- If more than one values are available for a given day *i*, X_i is the mean (or max, or min) of all the values of day *i*

Non Stationary – Absolute Time Encoding

A patient case for learning to predict the Magnesium value on Day 3 (after hospitalization)

Patient X	
Mg, Day 1	1.7
Mg, Day 2	Not Performed
Ca, Day 1	Not Performed
Ca, Day 2	9.0
Mg, Day 3	**Normal**



Matrix Encoding

	Mg_1	Mg_2	Ca_1	Ca_2	Mg_3
Patient X	1.7	Missing	Missing	9.0	**Normal**

All patients with a measurement on Day 3 for Mg can be used as examples to train a predictive model for Mg on Day 3

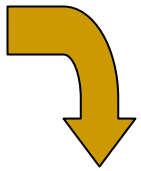
Non Stationary – Relative Time Encoding

- Define a time interval (e.g. a day)
- To create a training sample for predicting a lab on Day *K*, given *P* number of previous measurements
- For each temporal variable (predictor) *X (e.g.* Magnesium lab value)
  - define atemporal variables X_i denoting the value of *X*, *i* measurements before the time of prediction
  - define an atemporal variable TX_i denoting how long ago (from time of prediction) measurement X_i was taken

Non Stationary – Relative Time Encoding

A patient case for learning to predict the Magnesium value on Day 3 (after hospitalization) given 1 previous measurement of each lab

Patient X	
Mg, Day 1	1.7
Mg, Day 2	Not Performed
Ca, Day 1	Not Performed
Ca, Day 2	9.0
Mg, Day 3	**Normal**



Matrix Encoding

	Mg_1	TMg_1	Ca_1	TCa_1	Mg
Patient X	1.7	2 days	9.0	1 day	**Normal**

All patients with a measurement on Day 3 for Mg can be used as examples to train a predictive model for Mg on Day 3 using one prior measurement for all labs

Patient Data used for Training

- A single patient depending on the available data on her, may be used to train different models

Library of Models Approach for Predicting Modeling in Biomedicine

- TRAINING PHASE

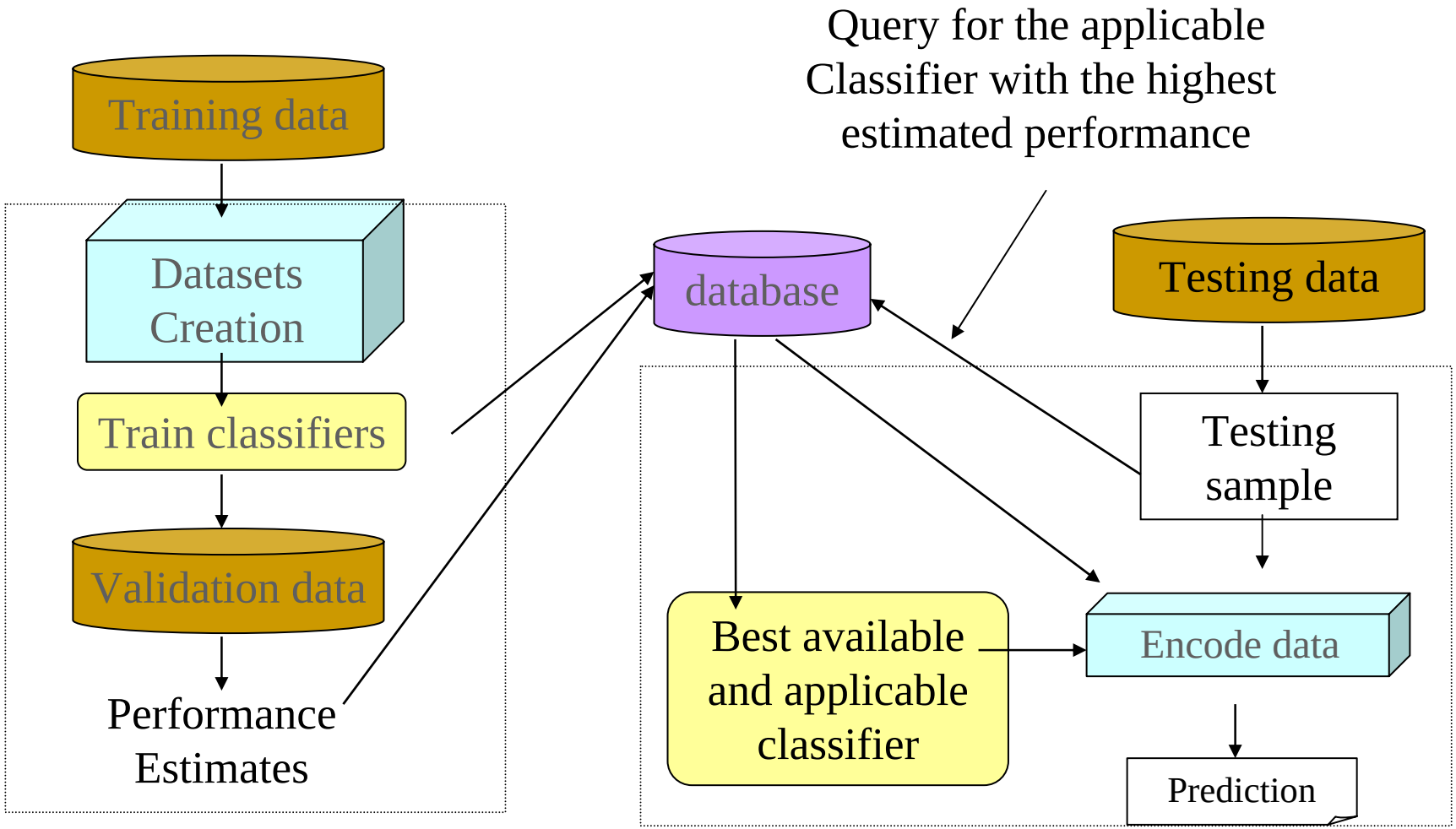
1. Break the population in (overlapping) sets that contain similar types and number of available predictors
2. Build a predictive model for each subpopulation
3. Estimate the performance of each model
4. Maintain all models in a Library of Models

Library of Models Approach for Predicting Modeling in Biomedicine

- To apply the Library of Models on a new laboratory test:
- APPLICATION PHASE

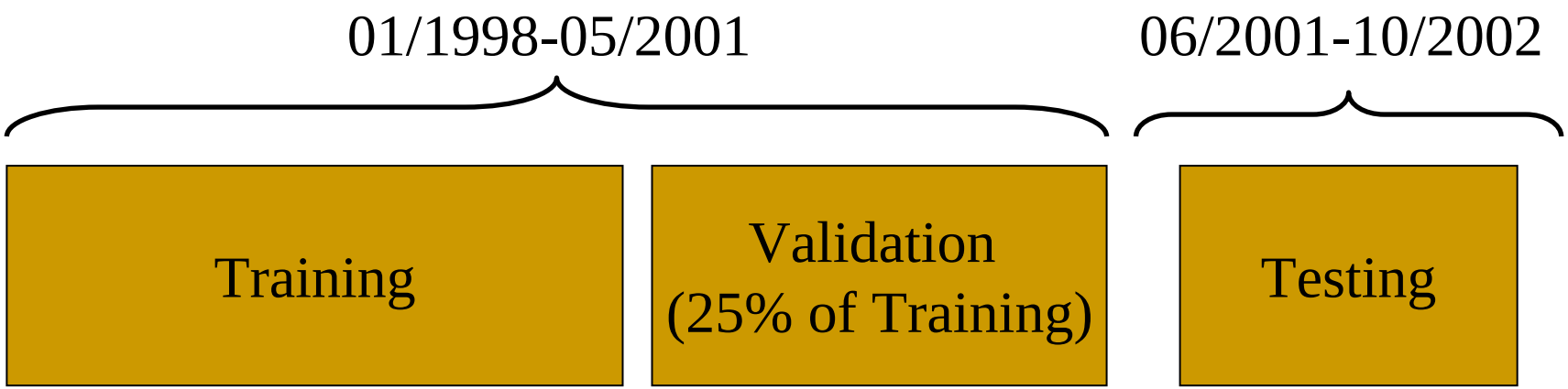
1. Find the applicable models in the library (depends on available information on the patient at that moment)
2. Select the model with the largest estimated performance
3. Encode the available information for use by the model
4. Apply the model to predict the result of the laboratory test

Library of Models Overview



Application of Library of Models to Prediction of Laboratory Abnormalities

Retrieved from StarPanel database ~$8 \cdot 10^6$ laboratory measurements on ~100,000 in-patients from Vanderbilt University Medical Center. Laboratory measurements were taken between 01/1998 and 10/2002.



Prediction Tasks

- Considered the following laboratory tests
  - (1) hematocrit, (2) blood urea nitrogen (BUN), (3) serum bicarbonate, (4) calcium, (5) creatinine, (6) ionized calcium, (7) magnesium, (8) osmolarity and (9) phosphorus
- Task: for laboratory *X* predict whether the next measurement on a given patient will fall within the range of "normal" values?
- Definition of normal
  - Find the values v_1 , $v_{2.5}$, $v_{97.5}$, v_{99} corresponding to the 1, 2.5, 97.5, 99 percentile of test *X* (as estimated by the training cases)
  - Six definitions of "normal" values
    - $[> v_1]$, $[v_1 , v_{99}]$, $[< v_{99}]$,
    - $[> v_{2.5}]$, $[v_{2.5} , v_{97.5}]$, $[< v_{97.5}]$

More on Encoding

- For each lab result used as a training example:
  - Target variable: 0/1 depending on whether the lab was within the normal range or not
  - Temporal variables encoded using Non-Stationary Absolute Time, Non-Stationary Relative Time, Stationary Absolute Time, Stationary Relative Time
  - (Binary) variables indicating the admission unit of the patient and the unit ordering a lab test for each lab used as a predictor
  - Cost and primary diagnosis from up to 2 prior hospitalizations
  - Demographic data
  - Indicator variables used to encode missing values

Dataset Creation

Create a different training dataset for each of the following options

- **T**: Laboratory test: {BUN, Ca, CaIo, CO2, Creat, Mg, Osmol, PCV, Phos}
- **R**: Range of normal values: {[> v_1], [v_1 , v_{99}], [< v_{99}], [> $v_{2.5}$], [$v_{2.5}$, $v_{97.5}$], [< $v_{97.5}$]
- **M**: Temporal modeling method: {SRT, SAT, NSRT, NSAT}
- **P**: Number of prior hospitalizations used: {0, 1, 2}
- **U**: Use variables corresponding to units where lab tests were ordered or not? {Yes, No}
- **K**: Number of previous measurements (or days) used in the model: {0, 1, 2, 3, 4, 5}
- **D**: Hospitalization day when the lab test to predict was taken (used only for NSRT and NSAT): {0, 1, 2, 3, 4, 5}

TOTAL: 16848 MODELS in the Library of Models, ~100K total models built, ~year single CPU time

Building the Predictive Models

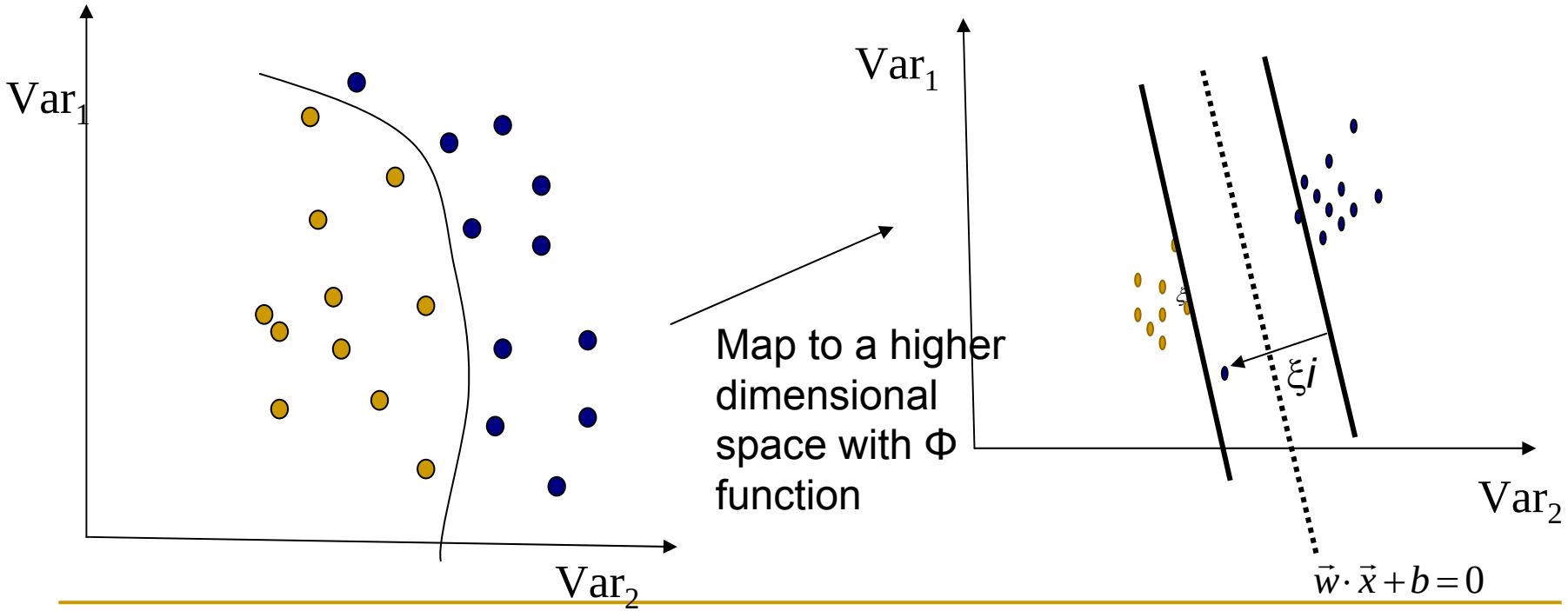
- Train a Support Vector Machine to learn a model from the data
  - Tune the parameters (kernel) using cross-validation
- Fit a sigmoid function to the decision surface of the model to convert the output of the Support Vector Machine to probabilities (e.g., prediction is normal with probability 70%)
- Estimate the performance of the model on the validation set using the Area Under the Receiver Operating Curve (AUC)

Support Vector Machines in one a Slide

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_i \xi_i$$

$$s.t.\ \ y_i(w \cdot \Phi(x) + b) \geq 1 - \xi_i, \forall x_i$$

$$\xi_i \geq 0$$



Classification Results

Area under ROC curve (without feature selection)

		Range of normal values					
		>1	**<99**	**[1, 99]**	**>2.5**	**<97.5**	**[2.5, 97.5]**
Laboratory test	**BUN**	75.9%	**93.4%**	68.5%	81.8%	**92.2%**	66.9%
	Ca	67.5%	80.4%	55.0%	77.4%	70.8%	60.0%
	Calo	63.5%	52.9%	58.8%	46.4%	66.3%	58.7%
	CO2	77.3%	**88.0%**	53.4%	77.5%	**90.5%**	58.1%
	Creat	62.2%	**88.4%**	83.5%	**88.4%**	**94.9%**	83.8%
	Mg	58.4%	71.8%	64.2%	67.0%	72.5%	62.1%
	Osmol	77.9%	64.8%	65.2%	79.2%	82.4%	71.5%
	PCV	62.3%	**91.6%**	69.7%	76.5%	84.6%	70.2%
	Phos	70.8%	75.4%	60.4%	68.0%	81.8%	65.9%

Classification Results

Excluding cases with no prior lab measurements

Area under ROC curve (without feature selection)

		Range of normal values					
		>1	**<99**	**[1, 99]**	**>2.5**	**<97.5**	**[2.5, 97.5]**
Laboratory test	**BUN**	80.4%	**99.1%**	76.6%	**87.1%**	**98.2%**	70.7%
	Ca	72.8%	**93.4%**	55.6%	81.4%	81.4%	63.4%
	Calo	74.1%	60.0%	50.1%	64.7%	72.3%	57.7%
	CO2	82.0%	**93.6%**	59.8%	84.4%	**94.5%**	56.3%
	Creat	62.8%	**97.7%**	**89.1%**	**91.5%**	**98.1%**	**87.7%**
	Mg	56.9%	70.0%	49.1%	58.6%	76.9%	59.1%
	Osmol	50.9%	60.8%	60.8%	**91.0%**	**90.5%**	68.0%
	PCV	74.9%	**99.2%**	66.3%	80.9%	80.6%	67.1%
	Phos	74.5%	**93.6%**	64.4%	71.7%	**92.2%**	69.7%

Improving Classification Using Feature Selection

- Number of features in datasets 100-10000
- Identify the smallest subset of predictors that exhibit the optimal prediction performance
- DSL-invented algorithms
  - HITON-PC, HITON-MB: causal interpretation of the selected features under certain conditions, optimality results
- Recursive Feature Elimination (linear and polynomial kernels used): Support-Vector Machine Based feature selection used in bioinformatics

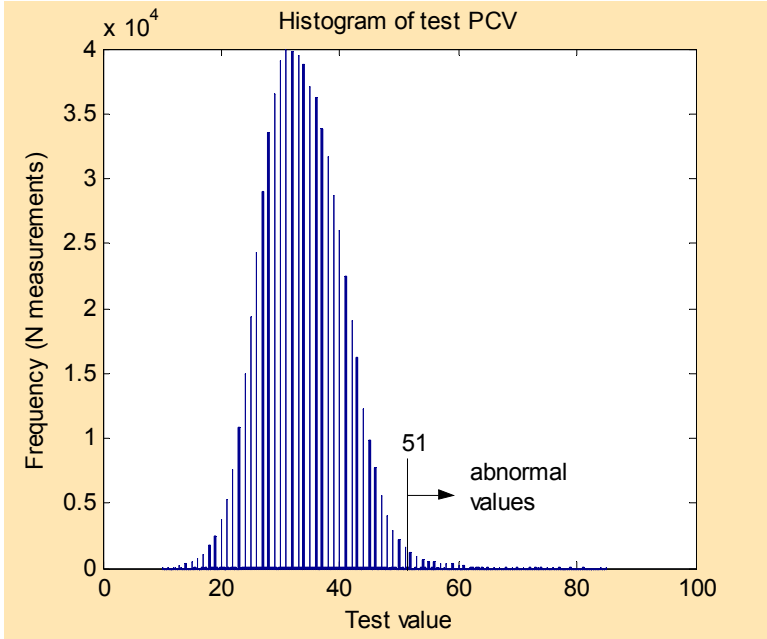
Improving Predictive Power and Parsimony of a PCV Model Using Feature Selection

Model description

Test name	PCV
Range of normal values	< 99 perc.
Data modeling	SAT
Number of previous measurements	1
Use variables corresponding to hospitalization units?	Yes
Number of prior hospitalizations used	0

Dataset description

	N samples (total)	N abnormal samples	N variables
Training set	10231	37	1124
Validation set	3411	13	
Testing set	1445	6	



Classification performance (area under ROC curve)

	All	RFE_Linear	RFE_Poly	HITON_PC	HITON_MB
Validation set	93.06%	90.91%	86.77%	90.23%	87.76%
Testing set	92.10%	79.37%	85.26%	98.06%	98.86%
Number of features	**1124**	**35**	**17**	**5**	**7**

Classification performance (area under ROC curve)

	All	RFE_Linear	RFE_Poly	HITON_PC	HITON_MB
Validation set	93.06%	90.91%	86.77%	90.23%	87.76%
Testing set	92.10%	79.37%	85.26%	98.06%	98.86%
Number of features	**1124**	**35**	**17**	**5**	**7**

Features

1	LAB: PM_1(Creat)	LAB: PM_1(PCV)	LAB: PM_1(PCV)	LAB: PM_1(PCV)
2	LAB: DT(PM_1(Creat))	DEMO: Gender	LAB: Indicator(PM_1(PCV))	LAB: PM_1(Calo)
3	LAB: PM_1(PCV)	LAB: Test Unit UNK (Test Mg, PM 1)	LAB: Test Unit UNK (Test BUN, PM 1)	LAB: Indicator(PM_1(PCV))
4	LAB: DT(PM_1(Gluc))	LAB: Test Unit NO_TEST_MEASUREMENT (Test Creat, PM 1)	LAB: Test Unit 4CNI (Test PCV, PM 1)	LAB: Indicator(PM_1(Calo))
5	LAB: PM_1(Ca)	LAB: Test Unit 4CNI (Test PCV, PM 1)	LAB: Test Unit NO_TEST_MEASUREMENT (Test Calo, PM 1)	LAB: DT(PM_1(Gluc))
6	LAB: Test Unit NO_TEST_MEASUREMENT (Test Creat, PM 1)	DEMO: Admission month		LAB: Test Unit UNK (Test BUN, PM 1)
7	LAB: Indicator(PM_1(Creat))	LAB: Test Unit NO_TEST_MEASUREMENT (Test Ca, PM 1)		LAB: Test Unit 4CNI (Test PCV, PM 1)
8	LAB: DT(PM_1(CO2))	LAB: Test Unit VHR (Test Calo, PM 1)		
9	LAB: DT(PM_1(Cl))	LAB: Test Unit NO_TEST_MEASUREMENT (Test K, PM 1)		
10	LAB: Test Unit 4CN (Test PCV, PM 1)	LAB: Test Unit 4CN (Test PCV, PM 1)		
11	LAB: Test Unit 3N/C (Test Phos, PM 1)	LAB: Test Unit NO_TEST_MEASUREMENT (Test Mg, PM 1)		
12	DEMO: Admission month	DEMO: Hospitalization Unit 4CNI		
13	LAB: DT(PM_1(Mg))	LAB: Test Unit NO_TEST_MEASUREMENT (Test Gluc, PM 1)		
14	LAB: Test Unit UNK (Test Mg, PM 1)	LAB: Test Unit NO_TEST_MEASUREMENT (Test Phos, PM 1)		
15	LAB: DT(PM_1(Ca))	LAB: Indicator(PM_1(Creat))		
16	LAB: DT(PM_1(Na))	LAB: Test Unit 4CNI (Test Mg, PM 1)		
17	LAB: PM_1(BUN)	LAB: Test Unit NO_TEST_MEASUREMENT (Test Na, PM 1)		
18	LAB: PM_1(Cl)			
19	LAB: Indicator(PM_1(Na))			
20	LAB: Test Unit NO_TEST_MEASUREMENT (Test Na, PM 1)			
21	DEMO: Hospitalization Unit 7SMI			
22	DEMO: Hospitalization Unit 4CN			
23	LAB: Test Unit 7SMI (Test PCV, PM 1)			
24	LAB: Test Unit 3N/C (Test K, PM 1)			
25	LAB: Test Unit NO_TEST_MEASUREMENT (Test Ca, PM 1)			
26	LAB: Indicator(PM_1(Ca))			
27	LAB: PM_1(K)			
28	LAB: Indicator(PM_1(CO2))			
29	LAB: Test Unit NO_TEST_MEASUREMENT (Test CO2, PM 1)			
30	LAB: DT(PM_1(Phos))			
31	LAB: DT(PM_1(K))			
32	LAB: PM_1(Calo)			
33	LAB: PM_1(Phos)			
34	LAB: Test Unit S44I (Test Phos, PM 1)			
35	LAB: Test Unit 3N/C (Test Calo, PM 1)			

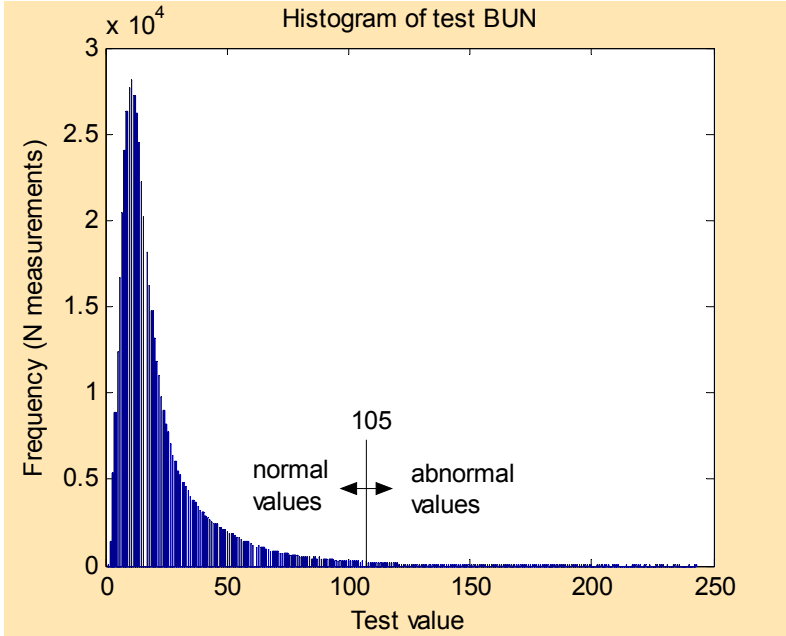
Improving Predictive Power and Parsimony of a BUN Model Using Feature Selection

Model description

Test name	BUN
Range of normal values	< 99 perc.
Data modeling	SRT
Number of previous measurements	5
Use variables corresponding to hospitalization units?	Yes
Number of prior hospitalizations used	2

Dataset description

	N samples (total)	N abnormal samples	N variables
Training set	3749	78	3442
Validation set	1251	27	
Testing set	836	16	



Classification performance (area under ROC curve)

	All	RFE_Linear	RFE_Poly	HITON_PC	HITON_MB
Validation set	95.29%	98.78%	98.76%	99.12%	98.90%
Testing set	94.72%	99.66%	99.63%	99.16%	99.05%
Number of features	**3442**	**26**	**3**	**11**	**17**

Classification performance (area under ROC curve)

	All	RFE_Linear	RFE_Poly	HITON_PC	HITON_MB
Validation set	95.29%	98.78%	98.76%	99.12%	98.90%
Testing set	94.72%	99.66%	99.63%	99.16%	99.05%
Number of features	**3442**	**26**	**3**	**11**	**17**

Features

1	LAB: PM_1(BUN)	LAB: PM_1(BUN)	LAB: PM_1(BUN)	LAB: PM_1(BUN)
2	LAB: PM_2(Cl)	LAB: Indicator(PM_1(Mg))	LAB: PM_5(Creat)	LAB: PM_5(Creat)
3	LAB: DT(PM_3(K))	LAB: Test Unit NO_TEST_MEASUREMENT (Test Calo, PM 1)	LAB: PM_1(Phos)	LAB: PM_3(PCV)
4	LAB: DT(PM_3(Creat))		LAB: Indicator(PM_1(BUN))	LAB: PM_1(Mg)
5	LAB: Test Unit J018 (Test Ca, PM 3)		LAB: Indicator(PM_5(Creat))	LAB: PM_1(Phos)
6	LAB: DT(PM_4(Cl))		LAB: Indicator(PM_1(Mg))	LAB: Indicator(PM_4(Creat))
7	LAB: DT(PM_3(Mg))		LAB: DT(PM_4(Creat))	LAB: Indicator(PM_5(Creat))
8	LAB: PM_1(Cl)		LAB: Test Unit 7SCC (Test Ca, PM 1)	LAB: Indicator(PM_3(PCV))
9	LAB: PM_3(Gluc)		LAB: Test Unit RADR (Test Ca, PM 5)	LAB: Indicator(PM_1(Phos))
10	LAB: DT(PM_1(CO2))		LAB: Test Unit 7SMI (Test PCV, PM 4)	LAB: DT(PM_4(Creat))
11	LAB: DT(PM_4(Gluc))		DEMO: Gender	LAB: Test Unit 11NM (Test BUN, PM 2)
12	LAB: PM_3(Mg)			LAB: Test Unit 7SCC (Test Ca, PM 1)
13	LAB: DT(PM_5(Mg))			LAB: Test Unit RADR (Test Ca, PM 5)
14	LAB: PM_1(PCV)			LAB: Test Unit 7SMI (Test PCV, PM 4)
15	LAB: PM_2(BUN)			LAB: Test Unit CCL (Test Phos, PM 1)
16	LAB: Test Unit 11NM (Test PCV, PM 2)			DEMO: Gender
17	LAB: Test Unit 7SCC (Test Mg, PM 3)			DEMO: Age
18	LAB: DT(PM_2(Phos))			
19	LAB: DT(PM_3(CO2))			
20	LAB: DT(PM_2(Gluc))			
21	LAB: DT(PM_5(Calo))			
22	DEMO: Hospitalization Unit TVOS			
23	LAB: PM_1(Phos)			
24	LAB: PM_2(Phos)			
25	LAB: Test Unit 11NM (Test K, PM 5)			
26	LAB: Test Unit VHR (Test Calo, PM 1)			

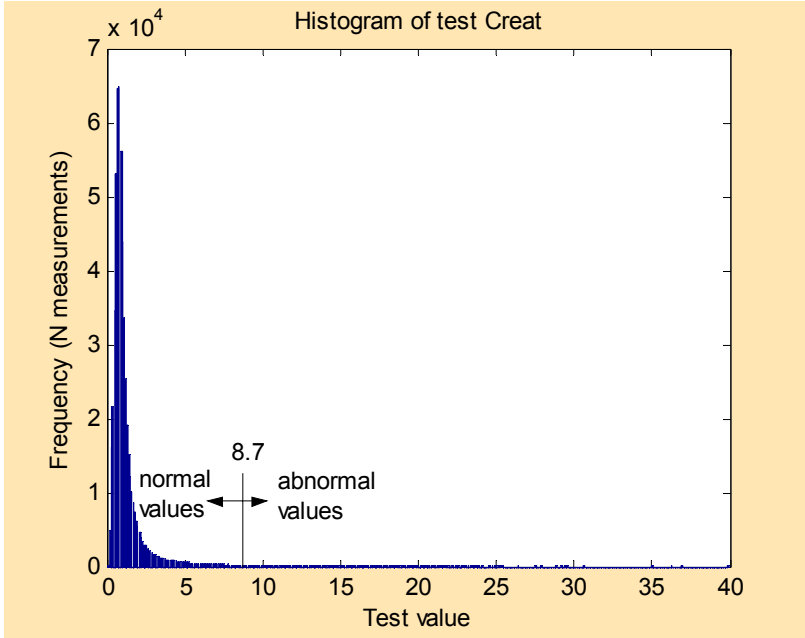
Improving Predictive Power and Parsimony of a Creatinine Model Using Feature Selection

Model description

Test name	Creat
Range of normal values	< 99 perc.
Data modeling	SRT
Number of previous measurements	3
Use variables corresponding to hospitalization units?	No
Number of prior hospitalizations used	0

Dataset description

	N samples (total)	N abnormal samples	N variables
Training set	4030	37	163
Validation set	1345	13	
Testing set	1710	7	



Classification performance (area under ROC curve)

	All	RFE_Linear	RFE_Poly	HITON_PC	HITON_MB
Validation set	98.88%	99.90%	99.79%	99.16%	99.10%
Testing set	93.39%	99.69%	99.72%	98.06%	97.26%
Number of features	**163**	**5**	**2**	**12**	**11**

Classification performance (area under ROC curve)

	All	RFE_Linear	RFE_Poly	HITON_PC	HITON_MB
Validation set	98.88%	99.90%	99.79%	99.16%	99.10%
Testing set	93.39%	99.69%	99.72%	98.06%	97.26%
Number of features	**163**	**5**	**2**	**12**	**11**

Features

1	LAB: PM_2(Creat)	LAB: Indicator(PM_2(Phos))	LAB: PM_1(CO2)	LAB: PM_1(CO2)
2	LAB: PM_1(Creat)	LAB: PM_1(Creat)	LAB: PM_3(Cl)	LAB: PM_1(Cl)
3	LAB: PM_3(Gluc)		LAB: PM_3(Creat)	LAB: PM_3(Cl)
4	LAB: DT(PM_2(PCV))		LAB: PM_1(Gluc)	LAB: PM_3(Creat)
5	LAB: PM_1(Gluc)		LAB: PM_2(PCV)	LAB: PM_1(Phos)
6			LAB: PM_2(Phos)	LAB: PM_2(Phos)
7			LAB: Indicator(PM_3(Cl))	LAB: Indicator(PM_1(Gluc))
8			LAB: Indicator(PM_3(Creat))	LAB: Indicator(PM_1(Phos))
9			LAB: Indicator(PM_1(Gluc))	LAB: Indicator(PM_2(Phos))
10			LAB: Indicator(PM_2(PCV))	DEMO: Hospitalization Unit 7N
11			LAB: Indicator(PM_2(Phos))	DEMO: Hospitalization Unit 9NSM
12			DEMO: Hospitalization Unit 9NSM	

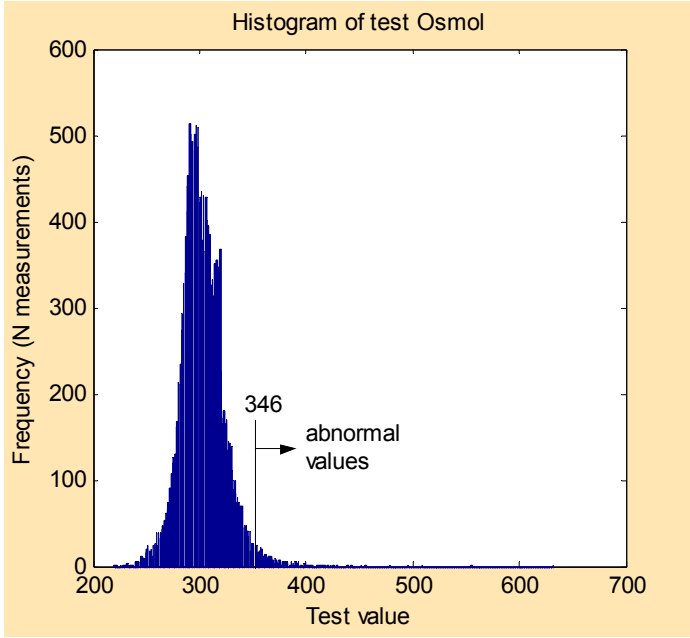
Improving Parsimony of an Osmolality Model Using Feature Selection

Model description

Test name	Osmol
Range of normal values	< 97.5 perc.
Data modeling	SAT
Number of previous measurements	1
Use variables corresponding to hospitalization units?	No
Number of prior hospitalizations used	1

Dataset description

	N samples (total)	N abnormal samples	N variables
Training set	2461	36	211
Validation set	821	12	
Testing set	1458	30	



Classification performance (area under ROC curve)

	All	RFE_Linear	RFE_Poly	HITON_PC	HITON_MB
Validation set	85.44%	86.58%	80.54%	81.08%	81.08%
Testing set	86.99%	79.82%	76.29%	83.16%	83.16%
Number of features	**211**	**26**	**6**	**3**	**3**

Classification performance (area under ROC curve)

	All	RFE_Linear	RFE_Poly	HITON_PC	HITON_MB
Validation set	85.44%	86.58%	80.54%	81.08%	81.08%
Testing set	86.99%	79.82%	76.29%	83.16%	83.16%
Number of features	**211**	**26**	**6**	**3**	**3**

Features

1	LAB: DT(PM_1(BUN))	LAB: PM_1(BUN)	LAB: PM_1(Osmol)	LAB: PM_1(Osmol)
2	LAB: Indicator(PM_1(Na))	LAB: PM_1(Creat)	LAB: Indicator(PM_1(Creat))	LAB: Indicator(PM_1(Creat))
3	LAB: PM_1(Cl)	LAB: PM_1(Cl)	LAB: Indicator(PM_1(Gluc))	LAB: Indicator(PM_1(Gluc))
4	LAB: PM_1(Creat)	LAB: PM_1(Gluc)		
5	LAB: DT(PM_1(Gluc))	LAB: PM_1(Na)		
6	LAB: PM_1(Gluc)	DEMO: Gender		
7	LAB: DT(PM_1(CO2))			
8	LAB: PM_1(K)			
9	LAB: PM_1(CO2)			
10	LAB: PM_1(Na)			
11	LAB: PM_1(Mg)			
12	LAB: PM_1(Osmol)			
13	LAB: Indicator(PM_1(Creat))			
14	LAB: Indicator(PM_1(BUN))			
15	LAB: Indicator(PM_1(Cl))			
16	LAB: Indicator(PM_1(Gluc))			
17	LAB: DT(PM_1(PCV))			
18	LAB: PM_1(Calo)			
19	LAB: PM_1(PCV)			
20	LAB: DT(PM_1(Na))			
21	LAB: DT(PM_1(K))			
22	LAB: PM_1(Ca)			
23	DEMOGRAPHICS: primary diagnosis 250 during the prior hospitalization 1			
24	DEMOGRAPHICS: renalchrg of prior hospitalization 1			
25	DEMOGRAPHICS: renalcost of prior hospitalization 1			
26	LAB: Indicator(PM_1(Calo))			

Conclusions

- Reducing inappropriate testing is an interesting and important problem
- First step in our approach is to learn to predict laboratory results
- Numerous predictive modeling and diagnostic problems in biomedicine contain data from heterogeneous populations in terms of available predictors
- Developed a new modeling methodology to address the problem, called Library of Models
- Applied it to a real clinical problem in a large scale preliminary study: the prediction of laboratory abnormalities
- The results of several tests can be reliably predicted

Future Directions

- Use clinical values for the definition of abnormalities
- Include more predictors, incorporate new data
- Employ methods developed for class-imbalanced data,
  - One class Support Vector Machines, SVMs that accept different misclassification costs for the two classes,
- Use boosting/n-fold cross-validation for estimation of performance
- Perform feature selection for all tasks
- Analyze the feature sets discovered from a clinical perspective
- Analysis of the prevalent encodings
- Better treatment of missing values
- Explore other temporal modeling methods

Future Directions

- Simulate the impact of the predictions to the clinical practice
  - How many tests can be saved for each misclassification of the system
- Intervention in a decision support system
- Evaluation