Aduenko A.A., Strijov V.V. Multimodelling and Object Selection for Banking Credit Scoring // Conference of the International Federation of Operational Research Societies, 2014 : 138.

4 - Technological Superiority
Jens Leth Hougaard, Mette Asmild

We develop a theoretical framework for analyzing technological possibilities. We consider fundamental properties of technology indexes and demonstrate that previous approaches violate a central axiom dubbed monotonicity in possibilities. From the axiomatic analysis emerge two canonical types of indexes: one based on the volume, and one based on the cardinality of the dominance set. We define a binary superiority relation where both types of indexes have to point in the same direction before concluding that one subset is superior to another.

■ TE-15

Tuesday, 16:00-17:30 - Room 125

Revenue Management with Advertising Applications

Stream: Revenue Management II
Invited session
Chair: *John Turner*

1 - Optimizing Online Advertising Budget Allocation across Multiple Placements
Jian Yang, Pengyuan Wang

Big online advertisers are typically faced with a challenging problem in campaign management: how to allocate advertising budget across multiple placements in order to maximize Return on Investment (ROI). We develop a Multi-Touch Attribution (MTA) methodology based on both observation and experimentation to measure ad effectiveness across multiple placements. The MTA empowers a simulator that provides advertisers with what-if analysis for budget allocation. We also build an optimization model using the MTA results to maximize the total ad effectiveness for advertisers, and hence their ROI.

2 - A Class of Nonlinear Allocation Problems with Heterogeneous Substitution
Huaxia Rui, De Liu, Andrew Whinston

We study the problem of efficiently allocating multiple types of goods (workloads) to multiple agents when different types of goods (workloads) are substitutable and the rates of substitutation differ across agents. We derive theoretical properties of such problems that enable us to design an extremely fast algorithm called SIMS for solving such problems. We expect the SIMS algorithm to work well for real-time applications with time-constrained allocation problems such as the allocation of online advertisement.

3 - The Least Cost Influence Problem
Rui Zhang, Dilek Gunnec, S. Raghavan

We analyze the diffusion process of a product over a social network while incentives are provided to the individuals. Such catalysation addresses the trade-off of minimizing the amount of incentives given and reaching a greater number of buyers. This problem is NP-Hard for general networks. However, we show that it is polynomially-solvable on tree networks under the assumption that all neighbors of a node exert equal influence. Next, we propose a totally unimodular integer programming formulation based on the insight that the influence propagation network must be a directed acyclic graph.

4 - Foundations of Social Network Ad Optimization
John Turner

We introduce revenue optimization models for placing ads in social networks, motivated by the connectivity structure of the underlying graph. We discuss some pros and cons of the underlying models, and illustrate our approach using real social graphs.

■ TE-16

Tuesday, 16:00-17:30 - Room 127

Model Selection Methods

Stream: Intelligent Optimization in Machine Learning and Data Analysis
Invited session
Chair: *Ivan Reyer*

1 - Multimodelling and Object Selection for Banking Credit Scoring
Alexander Aduenko, Vadim Strijov

To construct a bank credit scoring model one must select a set of informative objects (client records) to get the unbiased estimation of the model parameters. This set must have no outliers. The authors propose an object selection algorithm for mixture of regression models. It is based on analysis of the covariance matrix for the parameters estimations. The computational experiment shows statistical significance of the classification quality improvement. The algorithm is illustrated with the cash loans and heart disease data sets.

2 - Comparison of Different Clustering Algorithms Based PCF Classifiers
Emre Çimen, Gurkan Ozturk

In this study we dealt with generating different clustering algorithms based polyhedral conic classifiers. The main purpose of using clustering algorithms to generate PCF based classifiers is to determine the number of PCF's and divide the sets to the smaller parts. By this way stronger classifiers can be constructed. Expectation Maximization (EM) and k-Means based algorithms are implemented and tested on well-known literature test problems.

3 - Multicollinearity: Performance Analysis of Feature Selection Algorithms
Alexandr Katrutsa, Vadim Strijov

We investigate the multicollinearity problem and its influence on the performance of feature selection methods. The paper proposes the testing procedure for feature selection methods. We discuss the criteria for comparing feature selection methods according to their performance when the multicollinearity is present. Feature selection methods are compared according to the other evaluation measures. We propose the method of generating test data sets with different kinds of multicollinearity. Authors conclude about the performance of feature selection methods if the multicollinearity is present.

4 - Data Mining Application with Decision Tree Algorithms for the Evaluation of Personal Loan Customers' Repayment Performances
Aslı Çalış, Ahmet Boyacı, Kasım Baynal

Data mining techniques are used extensively in banking area such as many areas. In this study, conducted in banking sector, it was aimed to analysis of available personal loan customers and estimate potential customers' repayment performances with decision tree is one of the classification methods in data mining. In the study, SPSS Clementine was used as a software of data mining. An application was done with C5.0 and C&RT algorithms for evaluation of personal loan customers and the results were compared.

■ TE-17

Tuesday, 16:00-17:30 - Room 005

Conic Optimization and Applications

Stream: Interior Point Methods and Conic Optimization
Invited session
Chair: *Tamás Terlaky*

Multimodelling and Object Selection for Banking Credit Scoring

Alexander Aduenko

Moscow Institute of Physics and Technology
Department of Control and Applied Maths
Intellectual Systems Division

Mentor:
PhD in Maths, Vadim V. Strijov

Barcelona, 15th of July, 2014

Goals

Problems:

- Corruption and inconsistencies in credit scoring databases (outliers)
- Features multicollinearity
- Data non-uniformity

Goals:

- To design a method for outlier filtering in logistic regression
- To generalize the method for mixtures of logistic regression models and multilevel logistic models.

Designations

Regression model

$$f : \mathcal{X} \times \mathcal{W} \to \mathcal{Y}.$$

Regression function

$$f\,|_{\mathbf{w} \in \mathcal{W}} : \mathcal{X} \to \mathcal{Y}.$$

Data $\mathbf{x} \in \mathbb{R}^n$, $y \in \mathbb{R}$

$$D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}.$$

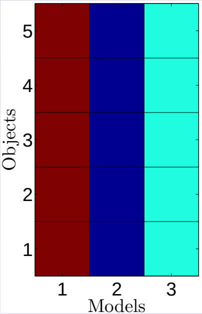
Feature matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$

$$\mathbf{X} = (\mathbf{x}_1^{\mathsf{T}}, \ldots, \mathbf{x}_m^{\mathsf{T}}).$$

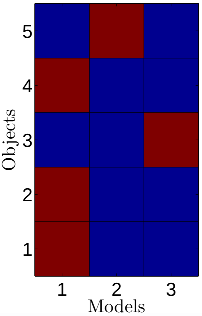
Mixture of models

Mixture of regression models — is the regression model $f = \sum_{k=1}^{K} \pi_k f_k(\mathbf{w}_k)$

where $\sum_{k=1}^{K} \pi_k = 1, \ \pi_k \geq 0.$



Multilevel regression model is a set of regression models f_k, $k = 1, \ldots, K$ such that the objects index set $\mathcal{I}$ is partitioned in subsets $\sqcup_{k=1}^{K} \mathcal{I}_k$ and for all the objects with indices from $\mathcal{I}_k$ the model f_k is used.



Basic hypothesis and assumptions

Basic assumptions

$$Y \sim Be(p),$$
$$p = f(\mathbf{x}^{\mathsf{T}}\mathbf{w}),$$
$$f(z) = \frac{1}{1+\exp(-z)}.$$

$$\mathsf{E}(Y_i) = p(\mathbf{x}) = f,$$
$$\mathsf{D}(Y_i) = p(\mathbf{x})(1-p(\mathbf{x})) = f(1-f).$$

Likelihood function

$$L(\mathbf{w}) = \prod_{i=1}^{m} f_i^{y_i}(1-f_i)^{1-y_i}.$$
$$l(\mathbf{w}) = -\sum_{i=1}^{m}\left(y_i \ln f_i + (1-y_i)\ln(1-f_i)\right).$$

Iterative parameter estimation

$$\mathbf{w}_j = \mathbf{w}_{j-1} - \mathbf{H}^{-1}(\mathbf{w}_{j-1})\nabla l(\mathbf{w}_{j-1}).$$

Newton-Rafson method for parameters estimation

$$\mathbf{w}_j = \mathbf{w}_{j-1} - (\mathbf{X}^{\mathsf{T}}\mathbf{R}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}(\mathbf{f}-\mathbf{y}),$$
$\nabla l(\mathbf{w}) = \mathbf{X}^{\mathsf{T}}(\mathbf{f}-\mathbf{y})$, $\mathbf{H} = \mathbf{X}^{\mathsf{T}}\mathbf{R}\mathbf{X}$, where
$$\mathbf{R} = \operatorname{diag}(\{f_i(1-f_i)\}_{i=1}^{m})$$

Estimation of covariance matrix for $\mathbf{w}$.

Using locally-normal approximation for a-posteriori $\mathbf{w}$ distribution, we get

$$\mathbf{w} \sim \mathcal{N}(\hat{\mathbf{w}},\ \mathbf{\Sigma}),$$

where by $\mathbf{\Sigma}$ denote the a-posteriori covariance matrix of the parameters.

Since $\nabla l(\hat{\mathbf{w}}) = 0$, using Taylor formula, obtain

$$\ln \frac{L(\mathbf{w})}{L(\hat{\mathbf{w}})} \approx -\frac{1}{2}(\mathbf{w} - \hat{\mathbf{w}})^{\mathsf{T}}\mathbf{H}(\mathbf{w} - \hat{\mathbf{w}}).$$

Finally we get

$$\mathbf{w} \sim \mathcal{N}(\hat{\mathbf{w}},\ \mathbf{H}^{-1}).$$

Object selection

Baseline methods

$$r_i^p = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1-\hat{p}_i)}}$$

$$r_i^d = \left\{2y_i \log\left(\frac{y_i}{\hat{p}_i}\right) + 2(1-y_i)\cdot \log\left(\frac{1-y_i}{1-\hat{p}_i}\right)\right\}^{1/2} \operatorname{sign}(y_i - \hat{p}_i)$$

Specificity definition

$\mathrm{Sp}(\mathbf{x}_i) = (\Delta_i \mathbf{w})^\mathsf{T} \mathbf{H} (\Delta_i \mathbf{w})$, where $\Delta_i \mathbf{w} = \hat{\mathbf{w}}_i - \hat{\mathbf{w}}$,

$\Delta_i \mathbf{w} \sim N(\mathbf{0}, \mathbf{H}^{-1})$.

$\mathrm{Sp}(\mathbf{x}_i) = (\Delta_i \mathbf{w})^\mathsf{T} \mathbf{H} (\Delta_i \mathbf{w}) \sim \sim \chi^2(|\mathcal{A}|)$.

Modification

A-priori: $\mathbf{w} \sim N(\mathbf{w}_0, \tau \mathbf{I})$

$$\mathbf{w} \sim N\left(\mathbf{w}_0, \left(\mathbf{H} + \frac{1}{\tau}\mathbf{I}\right)^{-1}\right)$$

$$\mathrm{Sp}_\tau(\mathbf{x}_i) = (\Delta_i \mathbf{w})^\mathsf{T} \left(\mathbf{H} + \frac{1}{\tau}\mathbf{I}\right) (\Delta_i \mathbf{w}).$$

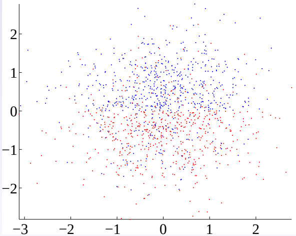
$\mathbf{H} \mapsto \operatorname{diag}(D_j)$, where

$$D_j = \frac{\sum_{i \in \mathcal{S}} (\Delta_i w_j)^2}{|\mathcal{S}| - 1}.$$

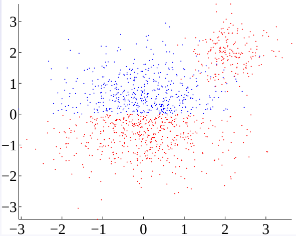
$$\mathrm{Sp}_w(\mathbf{x}_i, y_i) = \sum_{j=1}^{n} \frac{(\Delta_i w_j)^2}{D_j}$$

We used 4 benchmark data sets from UCI ML repository:.

- German cash loan: 1000 instances, 24 attributes, 2 classes
- Heart disease in South Africa: 462 instances, 13 attributes, 2 classes
- White wine quality: 4898 instances, 11 attributes and 2 classes
- Protein localization: 892 instances, 8 attributes and 2 classes

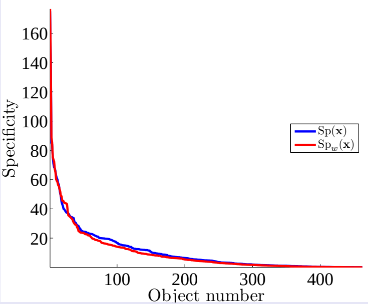


Non-clustered outliers

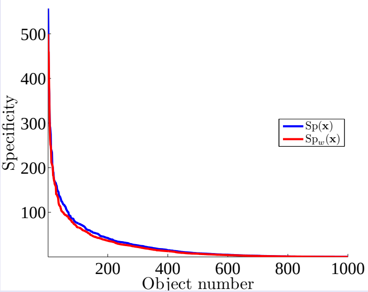


Clustered outliers

Comparison of two definitions for specificity



SAHD data set

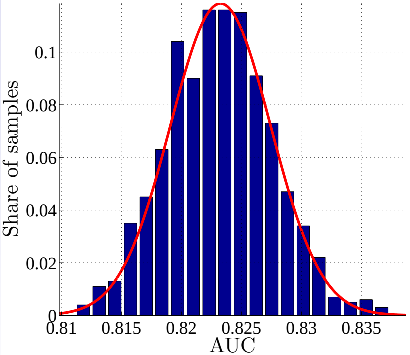


Cash loans data set

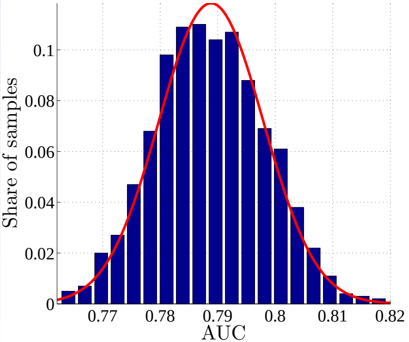
Data set \ Correlations	Pearson	Kendall
SAHD	0.9736	0.9132
Loans	0.9794	0.9377
Wine	0.9528	0.9028
Yeast	0.5230	0.8597

- Randomly generate many times a subset D_j of a sample set.
- Get the maximum likelihood estimates $\hat{\mathbf{w}}^j$ for the model parameters.
- Calculate the corresponding AUC value.

German cash loans data set



a) Learning sample



b) Testing sample

Checking significance of quality improvement

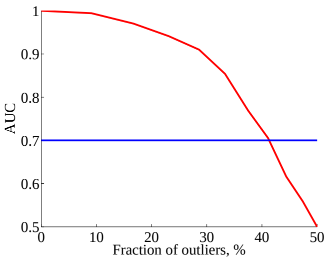
Properties \ Data	Loans	SAHD
AUC_{learn}	0.8819	0.8507
AUC_{test}	0.8308	0.8093
$\hat{m}$	0.8233; 0.7889	0.7994; 0.7722
$\hat{\sigma}$	0.0042; 0.0091	0.0061; 0.011
M	**14.0; 6.8**	**5.15; 3.32**
p_0	$\mathbf{0; 5.3 \cdot 10^{-12}}$	$\mathbf{1.3 \cdot 10^{-7}; 4.6 \cdot 10^{-4}}$

Properties \ Data	Wine	Yeast
AUC_{learn}	0.8109	0.7346
AUC_{test}	0.8084	0.7225
$\hat{m}$	0.7998; 0.7968	0.7142; 0.6965
$\hat{\sigma}$	0.0018; 0.0028	0.0049; 0.0076
M	**6.15; 4.15**	**4.18; 3.41**
p_0	$\mathbf{3.9 \cdot 10^{-10}; 1.7 \cdot 10^{-5}}$	$\mathbf{1.4 \cdot 10^{-5}; 3.2 \cdot 10^{-4}}$

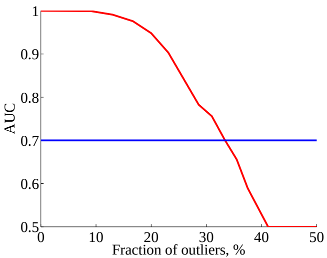
Performance on synthetic data sets having clustered and non-clustered outliers

Non-clustered outliers: $\mathbf{x} \in N(\mathbf{0}, \mathbf{I})$. $y_i = 1$ for $\mathbf{x}_i$ if $x_2 > 0$ and $y_i = 0$ otherwise. Outliers have the opposite rule to define the class label.

Clustered outliers: $\mathbf{x} \in N(\mathbf{0}, \mathbf{I})$. For non outliers $y_i = 1$ for $\mathbf{x}_i$ if $x_2 > 0$ and $y_i = 0$ otherwise. Outliers are generated from $N([2, 2]^\mathsf{T}, 0.5\mathbf{I})$. All outliers have the class label of 0.



Non-clustered outliers



Clustered outliers

Comparison to baseline methods

Data \ Methods	Pearson	Bayess	Spec.	t_p	t_b
SAHD	**0.7716**	0.7676	0.7661	-1.64	-0.45
Loans	**0.7868**	0.7864	0.7802	-2.71	-2.53
Wine	**0.7977**	0.7974	0.7970	-0.85	-0.42
Yeast	0.6845	**0.6951**	0.6944	5.88	-0.40
Non-clust., 9.1%	0.8997	**0.9021**	0.9002	0.25	-1.13
Non-clust.*, 9.1%	0.8945	0.8956	**0.8958**	0.80	0.16
Non-clust., 23.1%	0.7646	0.7653	**0.7665**	0.79	0.50
Non-clust.*, 23.1%	0.7671	0.7593	**0.7694**	0.99	4.33
Non-clust., 33.3%	0.6673	0.6679	**0.6680**	0.65	0.11
Non-clust.*, 33.3%	0.5372	0.6666	**0.6681**	64.6	0.75
Clustered, 9.1%	0.8885	0.9261	**0.9269**	20.9	0.44
Clustered*, 9.1%	0.8740	0.9515	**0.9541**	66.9	2.13
Clustered, 16.7%	0.8393	**0.8471**	0.8456	2.54	-0.63
Clustered*, 16.7%	0.8379	0.8305	**0.9060**	44.4	49.2
Clustered, 23.1%	0.8107	0.8171	**0.8174**	3.49	0.121
Clustered*, 23.1%	0.8105	0.7923	**0.8113**	0.28	6.53
Clustered, 33.3%	**0.7860**	0.7856	0.7853	-0.408	-0.18
Clustered*, 33.3%	0.7675	**0.7762**	0.7671	-0.108	-2.42

EM-algorithm for mixtures of models

Likelihood function for mixture of logistic models

$$L(\mathbf{w}_1, \ldots, \mathbf{w}_K, \boldsymbol{\pi}|\mathbf{X}, \mathbf{y}) =$$
$$= \prod_{i=1}^{m} \left(\sum_{k=1}^{K} \pi_k f(\mathbf{x}_i, \mathbf{w}_k)^{y_i} (1 - f(\mathbf{x}_i, \mathbf{w}_k))^{1-y_i} \right),$$
$$L(\mathbf{w}_1, \ldots, \mathbf{w}_K, , \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{y}) =$$
$$= \prod_{i=1}^{m} \prod_{k=1}^{K} \{\pi_k f(\mathbf{x}_i, \mathbf{w}_k)^{y_i} (1 - f(\mathbf{x}_i, \mathbf{w}_k))^{1-y_i}\}^{z_{ik}}.$$

E-step

$$\gamma_{ik} = \mathbb{E}[z_{ik}] = p(k|\mathbf{x}_i, \mathbf{w}_1, \ldots, \mathbf{w}_K, \boldsymbol{\pi}) =$$
$$= \frac{\pi_k f(\mathbf{x}_i, \mathbf{w}_k)^{y_i} (1 - f(\mathbf{x}_i, \mathbf{w}_k))^{1-y_i}}{\sum_{j=1}^{K} \pi_j f(\mathbf{x}_i, \mathbf{w}_j)^{y_i} (1 - f(\mathbf{x}_i, \mathbf{w}_j))^{1-y_i}},$$

M-step

$$\tilde{l}(\mathbf{w}_1, \ldots, \mathbf{w}_K, \boldsymbol{\pi}|\mathbf{X}, \mathbf{y}) = \mathbb{E}_{\mathbf{Z}}[-\log L(\mathbf{w}_1, \ldots, \mathbf{w}_K, \mathbf{Z}|\mathbf{X}, \mathbf{y})] =$$

$$= -\sum_{i=1}^{m}\sum_{k=1}^{K}\gamma_{ik}\{\log \pi_k + y_i \log(f(\mathbf{x}_i, \mathbf{w}_k)) + (1-y_i)\log(1 - f(\mathbf{x}_i, \mathbf{w}_k))\}.$$

$$\pi_k = \frac{1}{m}\sum_{i=1}^{m}\gamma_{ik}.$$

$$\tilde{l}(\mathbf{w}_1, \ldots, \mathbf{w}_K, \boldsymbol{\pi}|\mathbf{X}, \mathbf{y}) = -\sum_{k=1}^{K}\{\log \pi_k \sum_{i=1}^{m}\gamma_{ik}\} + \sum_{k=1}^{K}\tilde{l}_k(\mathbf{w}_k|\mathbf{X}, \mathbf{y}).$$

$$\frac{\partial \tilde{l}_k}{\partial \mathbf{w}_k} = \mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}_k(\mathbf{f} - \mathbf{y}),\ \boldsymbol{\Gamma}_k = \mathrm{diag}(\gamma_{ik}),$$

$$\mathbf{H}_k = \mathbf{X}^{\mathsf{T}}\mathbf{R}_k\mathbf{X},\ \mathbf{R}_k = \mathrm{diag}(\gamma_{ik} f(\mathbf{x}_i^{\mathsf{T}}\mathbf{w}_k) f(-\mathbf{x}_i^{\mathsf{T}}\mathbf{w}_k)).$$

1. Assume γ_{jk} are fixed.
2. Remove the object $\mathbf{x}_i$ from the sample and recalculate $\pi_1, \dots, \pi_K$
3. Reoptimize $\tilde{l}_k(\mathbf{w}_k|\mathbf{X}, \mathbf{y})$ across $\mathbf{w}_1, \dots, \mathbf{w}_K$
4. Define marginal specificity $\mathrm{Sp}_k(\mathbf{x}_i)$ for each model as $\mathrm{Sp}_k(\mathbf{x}_i) = (\mathbf{w}_k^{'} - \mathbf{w}_k)^{\mathsf{T}} \mathbf{H}_k^{-1} (\mathbf{w}_k^{'} - \mathbf{w}_k)$
5. Define integral specificity $\mathrm{Sp}(\mathbf{x}_i) = \sum\limits_{k=1}^{K} \mathrm{Sp}_k(\mathbf{x}_i)$

Notes:

$\nabla \tilde{l}_k(\mathbf{w}_k) = \sum\limits_{j=1,\, j \neq i}^{m} \gamma_{jk} \mathbf{x}_j (f_j - y_j) = -\gamma_{ik} \mathbf{x}_i (f_i - y_i) \Longrightarrow$ objects badly described by model $(f_i - y_i)$ or having high probability belonging to the model (γ_{ik}) have generally more influence on $\nabla \tilde{l}_k(\mathbf{w}_k)$.

- New method of object selection based on the introduced specificity function is proposed.
- For the common case of ill-conditioned hessian matrix the empiric specificity is proposed.
- High positive and monotonous correlation between specificity and empiric specificity is demonstrated.
- The method shows reasonable outliers discrimination for synthetic data sets having up to 40% of non-clustered outliers and up to 30% of clustered ouliers.
- Our results are significantly better than applying just the maximum likelihood estimator to the initial samples.
- Baseline algorithms show similar results for all considered benchmark data sets. Suggested method performs generally better for synthetic data sets having both clustered and non-clustered outliers.
- Generalization of the method for mixtures of logistic models is proposed.