Visual Numerics

Data Mining mit der JMSL™ Numerical Library for Java™ Applications

Stefan Sineux

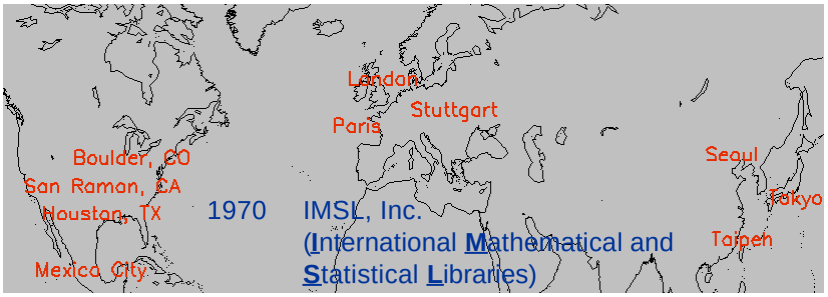
8. Java Forum Stuttgart

07.07.2005

Agenda

- Visual Numerics
- JMSL™ Numerical Library
- Neuronale Netze (Hintergrund)
- Demos
- Neuronale Netze (Details)
- Kundenbeispiele

Visual Numerics



1970 IMSL, Inc. (International Mathematical and Statistical Libraries)

1979 Precision Visuals

1992 Visual Numerics, Inc.

www.vni.com
Visual Numerics

Advanced Analytics

- Embeddable numerical analysis
  - Available for Java, .NET, C/C++, and Fortran
- Sophisticated Visual Data Analysis
  - Development environment including time-series and web-based functionality
- Expert Professional Services for customized algorithm and application development

Visual Numerics

www.vni.com

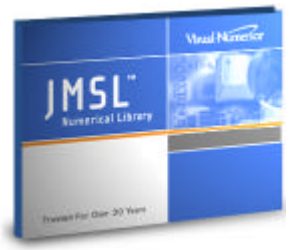
IMSL™ Family Benefits Overview

- Clean software architecture
  - Mainstream programming languages
- Accelerate development
- Develop better applications
- Develop flexible applications
- Improve quality and reduce uncertainty
- Reduce costs

Visual Numerics

www.vni.com

The JMSL™ Numerical Library



- Complete collection of mathematical, statistical and charting classes
- Written in 100% Java™, not an application or environment
  - Embeddable
  - Scalable
- Based upon the “gold standard” IMSL™ Library of algorithms
- Naturally suited for collaboration and sharing across the enterprise
- Fully supported and documented

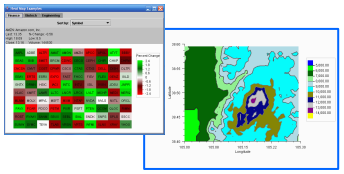
Visual Numerics

www.vni.com

Math, Stat, and Data Mining, Plus Charting

Mathematical, Statistical, Data Mining, and Finance

- Basic Types
- Linear Algebra
- Eigensystems
- Interpolation & Approximation
- Quadrature
- Differential Equations
- Nonlinear Equations
- Optimization
- Finance & Bond Calculations
- Variances, Covariances, & Correlations
- Time Series & Forecasting
- Multivariate Analysis
- Special Functions
- Basic Statistics
- Nonparametric Tests
- Regression
- Analysis of Variance
- Transforms
- Goodness of Fit
- Distribution Functions
- Random Number Generation
- Neural Network



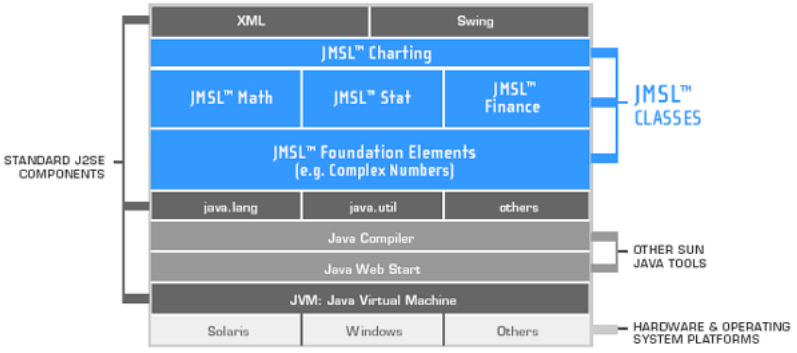
Charting

- Scatter
- Line
- High-Low -Close
- Pie
- Bar
- Histogram
- Log and Semilog
- Polar
- Area
- Contour
- Heat Map
- Box Plot
- Function and Spline
- Error Bar
- Support for XML
- Date/Time Support

Visual Numerics

www.vni.com

JMSL™ Architecture Overview



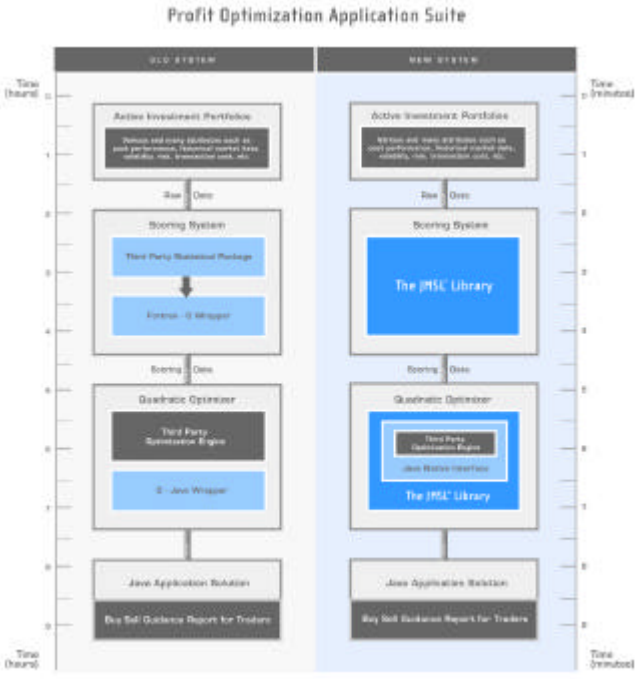
JMSL is the foundation for Java™ applications that require high-end math, statistical, and charting capabilities.

JMSL™ Implementation Options

- The JMSL Library is a pure Java™ class library. It can be used in any context in which Java J2SE can be used
  - Stand alone application running on a PC, Workstation, or Laptop
  - On a Web server
  - In an Applet within a browser
  - Using Java Web Start

Scalability and Embeddability using Numerics in Java™

- Clean Software Architecture
- Allows speed, simplicity, robustness
- Full native programming language
- Increased Performance
  - No wrappers that consume processor time
- Reduced complexity
- Faster Coding
- Easier Maintenance
- Increased stability



Visual Numerics

www.vni.com

What's New for the JMSL™ Numerical Library 3.0

- New Neural Network Data Mining Package

  Uniqueness:
  - Neural Network algorithms in 100% Java™
  - Fully embeddable technology
  - **Builds upon existing, comprehensive library with strong data mining and forecasting functionality**

  Key components:
  - Feed Forward Neural Network engine well suited for advanced predictive analysis
  - Neural Network pre- and post-processing algorithms for significant time savings
- Addition of important statistics algorithms for an even broader forecasting coverage
- New Chart Type
  - Heat Map to build upon strong, integrated Java charting package

Visual Numerics

www.vni.com

New Neural Network Data Mining Algorithms in 100% Java™

- Efficient and accurate predictive analysis on a highly scalable and collaborative environment
  - Data mining applications
  - Complex forecasting problems in corporate finance, insurance, business analytics, bioinformatics, life sciences
- Mimic human problem solving processes
  - Apply knowledge gained from past experience to new problems
  - Use historical data with known outcomes to build a predictive model of a complex system and use that data to train the network over time

Neural Network Technology: Main Categories of Applications

- **Forecasting:** predicting one or more quantitative outcomes from both quantitative and categorical input data,
- **Classification:** classifying input data into one of two or more categories, or
- **Statistical pattern recognition:** uncovering patterns, typically spatial or temporal, among a set of variables.

www.vni.com
Visual Numerics

Traditional Statistical Methods for Forecasting

- Forecasting and predictive models have been build using a variety of well known, traditional statistical methods
  - Linear regression and general least squares
  - Logistic regression and discrimination
  - Principal component analysis
  - Discriminant analysis
  - K-nearest neighbor classification
  - ARMA and ARIMA time series forecasts

Advantages of Neural Networks Over Traditional Methods (1/2)

- Neural Network configurations can be more simplified and yield the same solution as certain traditional statistical applications.
- Neural Networks provide a single framework for solving many traditional problems and can extend the range of problems that can be solved.

www.vni.com
Visual Numerics

Advantages of Neural Networks Over Traditional Methods (2/2)

- Neural Networks can provide more accurate and robust solutions for problems where traditional methods cannot be applied
  - Large amounts of noise
  - Short time series
  - Non stationary time series
  - Very large and complex systems
  - Apply without extensive analysis of underlying assumptions

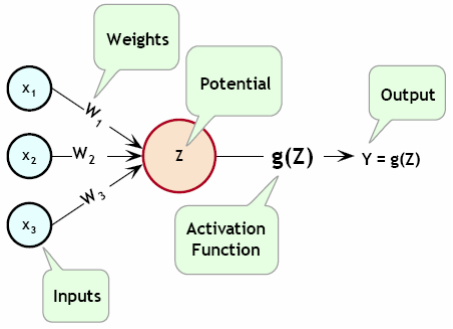
Visual Numerics

Demo

Visual Numerics

www.vni.com

Neural Network Terminology

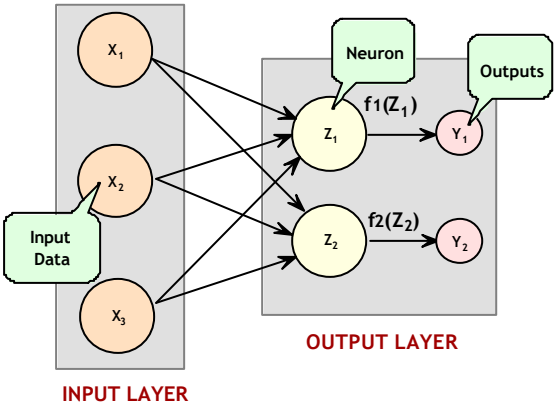


- *Weights* are applied to *inputs*
- An *activation function* is applied to the potential (*bias* term) at the *perceptron*
- An *output* is the result

Visual Numerics

www.vni.com

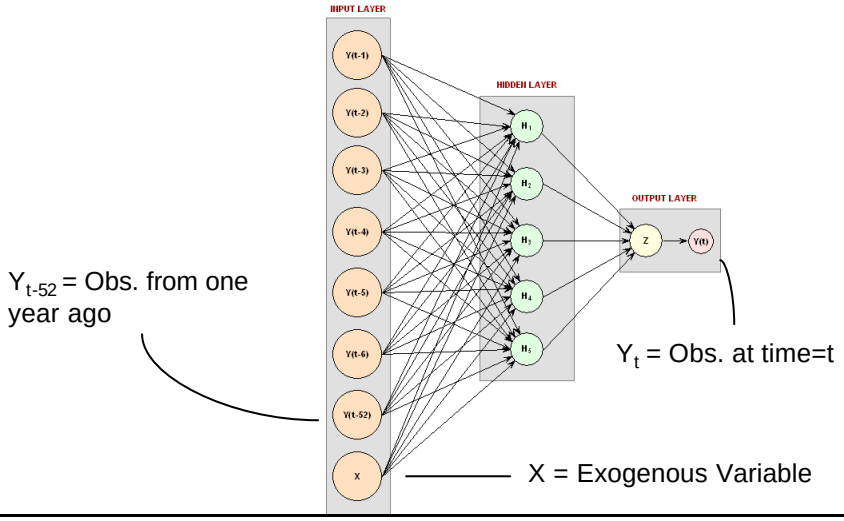
Single Layer Feed Forward Neural Network



Visual Numerics

www.vni.com

Time Series Forecasting with Exogenous Variables

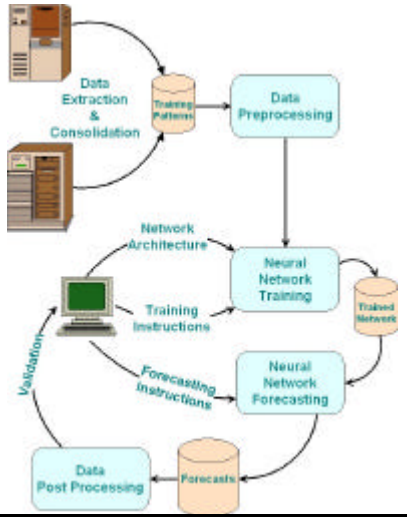


Y_{t-52} = Obs. from one year ago

Y_t = Obs. at time=t

X = Exogenous Variable

Neural Network Data Mining Process Flow Diagram



Visual Numerics components shown in blue rectangles

www.vni.com
Visual Numerics

Neural Network Data Mining Process Flow: Four Major Steps

- Data Filtering (“Data Pre-processing”)
  - Applied before sending data to a neural network training algorithm
  - Simplifies the network; makes training of the network more efficient
  - Continuous data, such as temperatures and pressures, should be scaled to fall within a range, normally –1 to 1
  - Reduce the dimensionality of data used as input to the network
  - Handling missing data
- Network Training
  - Use historical data to build a predictive model of a complex system
- Forecasting
  - Once the model is built, a current set of conditions can be fed to the model to generate a prediction of one or more target variables
- Network Validation
  - Compare known outcomes with forecasted outcomes
  - Train the network with a subset of the historical data

JMSL Classes and Methods
Pre- / postprocessing

- `ScaleFilter` *Scales or unscales continuous data prior to its use in neural network training, testing, or forecasting.*
- `TimeSeriesFilter` *Converts time series data to the format required for processing by a neural network.*
- `TimeSeriesClassFilter` *Converts time series data to the format required for processing by a neural network.*
- `UnsupervisedNominalFilter` *Converts nominal data into a series of binary encoded columns for input to a neural network. It also reverses the aforementioned encoding; accepting binary encoded data and returns a single nominal class.*
- `UnsupervisedOrdinalFilter` *Converts ordinal data into percentages. It also reverses encoding, accepting a percentage and converting it into an ordinal value.*

www.vni.com
Visual Numerics

JMSL Classes and Methods
Network Setup / Forecast

- `Network` *A neural network.*
- `FeedForwardNetwork` *A representation of a feed forward neural net network.*
- `Layer` *The base class for layers in a neural network.*
- `InputLayer` *Node layer in a neural net.*
- `HiddenLayer` *Hidden layer in a neural net.*
- `OutputLayer` *Output layer in a neural net.*
- `Node` *A node in a neural network.*
- `InputNode` *A node in the input layer.*
- `Perceptron` *A Perceptron node in a neural network. Perceptrons are created by a factory method in a layer.*
- `OutputPerceptron` *A Perceptron in the output layer.*
- `Link` *A link in the neural network.*

JMSL Classes and Methods Training

- `QuasiNewtonTrainer` *Trains a network using the quasi-Newton method,* `MinUnconMultiVar`.
- `LeastSquaresTrainer` *Trains a feed forward engine using nonlinear least squares solver.*
- `EpochTrainer` *Train the network with randomly selected subsets of the training patterns.*

www.vni.com
Visual Numerics

Teksouth

Financial Application

Need **Predictive Model to Forecast Compensation Expenses**

Solution **Custom Built Neural Net Analytics Application**

- Ability to better project staffing needs and accurately predict budget expenses
- Data mining, learning engine continuously refines forecasting model
- Ability for military agencies to much more effectively manage budgetary process
- Full application with easy ability to share across the enterprise

TEKSOUTH
CORPORATION

European Grocery Store Chain

- Problem
  - Inventory Planning and Forecasting
  - 200,000+ products and multiple stores
  - Previously relied on historical averages for forecasting
  - Unable to forecast short-term inventory requirements
- Goal
  - Reduce inventory costs through accurate, short-term forecasting of demand by product, by store

Visual Numerics
www.vni.com

European Grocery Store Chain

- Total Data Mining Solution
  - Numerical Libraries
  - Data Mining expertise
  - Packaged as a customized, configurable, easy-to-use solution
- Technology
  - Neural Net preprocessing
  - Forecasting and Training engines
  - Integrated into customers' intranet applications used by store managers

Summary

- Forecasting techniques can bring real competitive value to many organizations.
- There are a wide range of techniques available for forecasting and Visual Numerics has the tools and expertise to deliver the right solution for any application.

www.vni.com
Visual Numerics

Major European Wireless Provider

Telecommunications Industry

Need **Forecasting for Cellular Network Optimization**

Solution **IMSL™ Auto_ARIMA Algorithm**

- Required expert algorithm development to combine the concepts of two recent published papers on statistical methods
- Automatically identify and classify outliers (spikes in the data)
- Automatically locate any seasonal trends within the data
- Adjust the forecast to minimize the effect of any outliers and seasonality issues
- Automatically determine the best (p,d,q) ARIMA model for a particular dataset
- Provide the ability to pass a large number of datasets into the algorithm in a timely manner

Visual Numerics

Q&A

Visual Numerics

www.vni.com

Weitere Informationen

Wir würden uns freuen, Sie an
Stand Nr. 30 begrüßen zu dürfen!

Kontakt:

Visual Numerics International GmbH
Curiestr. 2
70563 Stuttgart

Tel. +49 711 / 67400-260
Fax +49 711 / 67400-456
Email info@visual-numerics.de
Web http://www.visual-numerics.de