N479: Applied Statistical Modeling and Big Data Analytics

Instructor(s): Srikanta Mishra

Format and Duration
Classroom - 3 Days
Virtual - 5 Sessions

Summary

This training course offers a hands-on introduction to statistical modeling and big data analytics specifically designed for petroleum engineering and geoscience applications. Participants will engage with a variety of topics, including: (a) easy-to-understand descriptions of the commonly-used statistical techniques, (b) case studies demonstrating the applicability, limitations and value-added proposition for these methods, and (c) hands-on problems sessions using open source and/or commercial software.

Business Impact: By the end of this course, engineers and geologists will be equipped with practical techniques for identifying hidden patterns and relationships within large datasets, enabling them to derive actionable, data-driven insights. Attendees will gain advanced capabilities to extract new insights from E&P data that can help: (a) learn hidden patterns and relationships in geologic datasets, (b) identify production sweet spots in developed plays; (c) determine factors responsible for separating good wells from poor producers wells, (d) build fast surrogate models of reservoir performance, and (e) assist in predictive maintenance by identifying failure inducing conditions from historical records.

Learning Outcomes

Participants will learn to:

1. Apply foundational concepts in probability and statistics for basic data analysis
2. Perform linear regression for building simple input-output models
3. Conduct multivariate data reduction and clustering for finding sub-groups of data that have similar attributes
4. Converse with confidence about big data, data analytics and machine learning terminology and fundamental concepts, and critically review topical technical publications
5. Apply machine learning techniques for regression and classification for developing data-driven input-output models
6. Evaluate proxy modeling and uncertainty quantification studies for probabilistic performance forecasting

Training Method

This is a classroom or virtual classroom consisting of lectures interspersed with worked examples, hands-on exercises and discussions.

The textbook for the course will be "Applied Statistical Modeling and Data Analytics: A Practical Guide for the Petroleum Geosciences" by Srikanta Mishra and Akhil Datta-Gupta (Elsevier, 2017), supplemented by various technical publications.

Who Should Attend

This course is for designed for petroleum engineers, geoscientists, and managers interested in becoming

N479: Applied Statistical Modeling and Big Data Analytics

Instructor(s): Srikanta Mishra

Format and Duration

Classroom - 3 Days
Virtual - 5 Sessions

smart users of statistical modeling and data analytics.

Course Content

Part 1

- Foundational Concepts
  - Big data technologies, basic data analytics and machine learning terminology/concepts
  - Data, statistics, and probability
  - Distributions (models, fitting distributions to data)
  - Inference (Confidence limits, bootstrap, significance tests)
  - Data visualization
  - Problem session
- Basic Regression Analysis
  - Linear regression (univariate and multivariate regression)
  - Understanding regression statistics, ANOVA
  - Non-parametric regression
  - Problem session

Part 2

- Multivariate Data Analysis
  - Dimension reduction (Principal component analysis)
  - Cluster analysis (K-means, hierarchical clustering, self-organizing maps)
  - Problem session
- Machine Learning Basics
  - Overview of techniques
  - Evaluating model performance (model validation, goodness-of-fit, common pitfalls)
  - Variable importance
  - Model aggregation
  - Case studies

Part 3

- Machine Learning for Regression and Classification I
  - Classification/regression trees
  - Random forest
  - Gradient boosting machine
  - Problem session

Part 4

- Machine Learning for Regression and Classification II

- Support vector machine
- Neural networks and deep learning
- Problem session

Part 5

- Miscellaneous topics and Wrap-up
  - Experimental design and response surface analysis
  - Uncertainty quantification
  - Selected literature review
  - Key takeaways and resources
  - Data analytics do's and don't's