Machine Learning Internals

Course Summary

Description

This training course provides a deep dive into machine learning, datamining, and statistical pattern recognition.

The course will also draw from numerous case studies and applications, so that you'll also learn how to apply learning algorithms to building smart robots (perception, control), text understanding (web search, anti-spam), computer vision, medical informatics, audio, database mining, and other areas.

Topics

- Introduction
- Machine Learning for Predictive Analytics
- Data to Insights to Decisions
- Data Exploration
- Advanced Data Exploration
- Information-based Learning
- Similarity-based Learning
- Probability-based Learning
- Error-based Learning
- Evaluation
- Software Tools
- Linear Regression
- Logistic Regression
- Generalized Linear Models and the Exponential Family
- Directed Graphical Models (Bayes Nets)
- Latent Linear Models
- Kernels
- Clustering
- Introduction to Deep Learning

Audience

This course is designed for those seeking a deep dive into machine learning, datamining, and statistical pattern recognition.

Prerequisites

There are no prerequisites for this course.

Duration

Five days

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically

Machine Learning Internals

Course Outline

I. Introduction
- A. Model selection
- B. Supervised learning
- C. Discovering graph structure
- D. Types of machine learning
- E. Machine learning: what and why?
- F. Parametric vs non-parametric models
- G. No free lunch theorem
- H. Linear regression
- I. Some basic concepts in machine learning
- J. Discovering clusters
- K. Classification
- L. Regression
- M. Matrix completion
- N. Logistic regression
- O. Parametric models for classification and regression
- P. The curse of dimensionality
- Q. Overfitting
- R. Unsupervised learning
- S. Discovering latent factors
- T. A simple non-parametric classifier: K-nearest neighbors

II. Machine Learning for Predictive Analytics
- A. Predictive Data Analytics Tools
- B. How Does Machine Learning Work?
- C. The Road Ahead
- D. What Can Go Wrong with Machine Learning?
- E. The Predictive Data Analytics Project Lifecycle: CRISP-DM
- F. What Is Machine Learning?
- G. What Is Predictive Data Analytics?

III. Data to Insights to Decisions
- A. Different Types of Data
- B. Different Types of Features
- C. Designing the Analytics Base Table
- D. Designing and Implementing Features
- E. Assessing Feasibility
- F. Converting Business Problems into Analytics Solutions
- G. Case Study: Motor Insurance Fraudmotor
- H. Implementing Features
- I. Handling Time

IV. Data Exploration
- A. Outliers
- B. Handling Missing Values
- C. Handling Outliers
- D. Missing Values
- E. Irregular Cardinality
- F. Handling Data Quality Issues
- G. The Data Quality Report
- H. The Normal Distribution
- I. Identifying Data Quality Issues
- J. Getting to Know the Data

V. Advanced Data Exploration
- A. Measuring Covariance and Correlation
- B. Visualizing Relationships Between Features
- C. Binning
- D. Data Preparation
- E. Normalization

VI. Information-based Learning
- A. Shannon's Entropy Model
- B. Handling Continuous Descriptive Features
- C. Decision Trees
- D. Predicting Continuous Targets
- E. Extensions and Variations
- F. Fundamentals
- G. Information Gain
- H. Big Idea
- I. Standard Approach: The ID Algorithm
- J. Tree Pruning
- K. Alternative Feature Selection and Impurity Metrics

VII. Similarity-based Learning
- A. Standard Approach: The Nearest Neighbor Algorithm
- B. Predicting Continuous Targets
- C. Fundamentals
- D. Other Measures of Similarity
- E. Extensions and Variations
- F. Data Normalization
- G. Feature Space
- H. Big Idea
- I. Measuring Similarity Using Distance Metrics
- J. Feature Selection
- K. Handling Noisy Data
- L. Efficient Memory Search

VIII. Probability-based Learning
- A. Big Idea
- B. Smoothing
- C. Extensions and Variations

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically

Machine Learning Internals

Course Outline (cont'd)

D. Bayes' Theorem
E. Bayesian Networks
F. Continuous Features: Probability Density Functions
G. Continuous Features: Binning
H. Bayesian Prediction
I. Conditional Independence and Factorization
J. Fundamentals
K. Standard Approach: The Naive Bayes Model

IX. Error-based Learning
A. Setting the Learning Rate Using Weight Decay
B. Error Surfaces
C. Multinomial Logistic Regression
D. Modeling Non-linear Relationships
E. Handling Categorical Descriptive Features
F. Interpreting Multivariable Linear Regression Models
G. Simple Linear Regression
H. Big Idea
I. Handling Categorical Target Features: Logistic Regression
J. Extensions and Variations
K. Fundamentals
L. Choosing Learning Rates and Initial Weights
M. Standard Approach: Multivariable Linear Regression with Gradient
N. Descent
O. Gradient Descent
P. Multivariable Linear Regression
Q. Measuring Error

X. Evaluation
A. Performance Measures: Prediction Scores
B. Designing Evaluation Experiments
C. Evaluating Models after Deployment
D. Performance Measures: Multinomial Targets
E. Extensions and Variations
F. Fundamentals
G. Performance Measures: Continuous Targets
H. Performance Measures: Categorical Targets
I. Big Idea
J. Standard Approach: Misclassification Rate on a Hold-out Test Set

XI. Software Tools:
A. Matlab
B. H2O
C. Spark ML/Mlib
D. Octave

XII. Linear Regression:
A. Regularization effects of big data
B. Bayesian inference when ?^2 is unknown
C. Model specification
D. Numerically stable computation
E. Computing the posterior
F. Geometric interpretation
G. Convexity
H. Connection with PCA
I. Maximum likelihood estimation (least squares)
J. Bayesian linear regression
K. Derivation of the MLE
L. Computing the posterior predictive
M. EB for linear regression (evidence procedure)
N. Ridge regression
O. Basic idea
P. Robust linear regression
Q. Introduction

XIII. Logistic Regression
A. Residual analysis (outlier detection)
B. Generative vs discriminative classifier
C. Multi-class logistic regression
D. Online learning and regret minimization
E. Iteratively reweighted least squares (IRLS)
F. Quasi-Newton (variable metric) methods
G. Newton's method
H. Bayesian logistic regression
I. A Bayesian view
J. Laplace approximation
K. l2 regularization
L. Gaussian approximation for logistic regression
M. Approximating the posterior predictive
N. Derivation of the BIC
O. Steepest descent
P. Introduction
Q. MLE
R. Model specification

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically

Machine Learning Internals

Course Outline (cont'd)

S. Online learning and stochastic optimization
T. Dealing with missing data
U. Fisher's linear discriminant analysis (FLDA)
V. Model fitting
W. Stochastic optimization and risk minimization
X. Pros and cons of each approach
Y. The LMS algorithm
Z. Logistic regression
AA. The perceptron algorithm

XIV. Generalized Linear Models and the Exponential Family
A. Introduction
B. Maximum entropy derivation of the exponential family
C. Ordinal probit regression
D. Generalized linear mixed models
E. Examples
F. semi-parametric GLMMs for medical data
G. The pointwise approach
H. Computational issues
I. Application to domain adaptation
J. ML and MAP estimation
K. Learning to rank
L. ML/MAP estimation using gradient-based optimization
M. Multinomial probit models
N. Probit regression
O. Bayesian inference
P. Generalized linear models (GLMs)
Q. Other kinds of prior
R. Basics
S. The exponential family
T. The pairwise approach
U. Log partition function
V. Loss functions for ranking
W. Bayes for the exponential family
X. Definition
Y. Hierarchical Bayes for multi-task learning
Z. Application to personalized email spam filtering
AA. The listwise approach
BB. Latent variable interpretation
CC. MLE for the exponential family
DD. Multi-task learning

XV. Directed Graphical Models (Bayes Nets)
A. Chain rule
B. Markov and hidden Markov models
C. Introduction
D. Graph terminology
E. Naive Bayes classifiers
F. d-separation and the Bayes Ball algorithm (global Markov properties)
G. Learning with missing and/or latent variables
H. Conditional independence
I. Other Markov properties of DGMs
J. Inference
K. Genetic linkage analysis
L. Directed Gaussian graphical models
M. Influence (decision) diagrams
N. Learning
O. Graphical models
P. Directed graphical models
Q. Markov blanket and full conditionals
R. Plate notation
S. Learning from complete data
T. Conditional independence properties of DGMs

XVI. Latent Linear Models
A. Other estimation principles
B. Principal components analysis (PCA)
C. Probabilistic PCA
D. Using EM
E. The FastICA algorithm
F. FA is a low rank parameterization of an MVN
G. Fitting FA models with missing data
H. Unidentifiability
I. Choosing the number of latent dimensions
J. Partial least squares
K. Singular value decomposition (SVD)
L. EM for factor analysis models
M. Mixtures of factor analysers
N. EM algorithm for PCA
O. Model selection for FA/PPCA
P. Supervised PCA (latent factor regression)
Q. Canonical correlation analysis
R. PCA for categorical data
S. Maximum likelihood estimation
T. PCA for paired and multi-view data
U. Inference of the latent factors
V. Classical PCA: statement of the theorem

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically

Machine Learning Internals

Course Outline (cont'd)

W. Model selection for PCA
X. Independent Component Analysis (ICA)
Y. Factor analysis

XVII. Kernels
A. Smoothing kernels
B. Kernels for comparing documents
C. The kernel trick
D. SVMs for classification
E. Kernelized ridge regression
F. SVMs for regression
G. Linear kernels
H. Kernel machines
I. Introduction
J. Comparison of discriminative kernel methods
K. Kernelized nearest neighbor classification
L. Using kernels inside GLMs
M. A probabilistic interpretation of SVMs
N. Kernel functions
O. RBF kernels
P. Mercer (positive definite) kernels
Q. Kernel density estimation (KDE)
R. Kernel PCA
S. String kernels
T. LVMs, RVMs, and other sparse vector machines
U. Kernels for building generative models
V. Choosing C
W. Kernelized K-medoids clustering
X. Pyramid match kernels
Y. Kernel regression
Z. Kernels derived from probabilistic generative models
AA. Locally weighted regression
BB. Summary of key points
CC. Support vector machines (SVMs)
DD. From KDE to KNN
EE. Matern kernels

XVIII. Clustering
A. Agglomerative clustering
B. The Dirichlet process
C. Clustering datapoints and features
D. Graph Laplacian
E. Evaluating the output of clustering methods
F. Dirichlet process mixture models
G. Multi-view clustering
H. Spectral clustering
I. Applying Dirichlet processes to mixture modeling
J. Bi-clustering
K. Fitting a DP mixture model
L. Choosing the number of clusters
M. Measuring (dis)similarity
N. Affinity propagation
O. Introduction
P. From finite to infinite mixture models
Q. Bayesian hierarchical clustering
R. Normalized graph Laplacian
S. Hierarchical clustering
T. Divisive clustering

XIX. Introduction to Deep Learning
A. Learning image features using d convolutional DBNs
B. Deep generative models
C. Information retrieval using deep auto-encoders (semantic hashing)
D. Deep directed networks
E. Data visualization and feature discovery using deep auto-encoders
F. Deep Boltzmann machines
G. Applications of deep networks
H. Stacked de-noising auto-encoders
I. Learning audio features using d convolutional DBNs
J. Greedy layer-wise learning of DBNs
K. Deep neural networks
L. Deep belief networks
M. Deep multi-layer perceptrons
N. Deep auto-encoders
O. Introduction
P. Handwritten digit classification using DBNs

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically