Enhance IT

Data Science Course Agenda

Week 1: Supervised Machine Learning: Linear Regression

- Simple Linear Regression
- Multiple Linear Regression
- Regression via Linear Combination of Basis Functions
- Multivariate Linear Regression
- The Ordinary Least Squares Objective Function
- The Probabilistic Interpretation of OLS (MLE)
- Ridge Regression (L^2 Regularization)
- The Probabilistic Interpretation of Ridge Regression (MaP)
- Nuisance Correlation and LASSO
- Regression (L^1 Regularization)
- The Probabilistic Interpretation of LASSO Regression (MaP)
- ElasticNet Regression
- The Normal Equation Solution
- Gradient Descent
- Measuring Goodness of Fit (R-squared)
- Overfitting
- Interpreting the Model
- Cross Validation and Model Selection
- How to know what to try next after fitting a linear model
- Bias vs. Variance
- Diagnosing Bais vs. Variance Issues in Higher Dimensional Feature Spaces

Week 2: Supervised Machine Learning: Logistic Regression

- Linear Classification
- The Logistic Function – Sigmoid
- Binary Logistic Regression

- Softmax
- Multinomial Logistic Regression
- The Binary Cross Entropy Objective Function
- The General Cross Entropy Objective Function
- The Probabilistic Interpretation of Cross Entropy
- Ridge Regression (L^2 Regularization)
- Nuisance Corrolation and LASSO Regression (L^1 Regularization)
- ElasticNet Regression
- Gradient Descent for Logistic Regression Measuring Goodness of Fit:
    - Classification Rate Precision
    - Recall
    - F Score
    - The Confusion Matrix
    - Receiving Operator Characteristic and Area Under the Curve

Week 3 Supervised Machine Learning: Artificial Neural Networks

- Artificial Neural Networks for Classification Artificial Neural Networks for Regression
- Forward Propagation
- Common Activation Functions and Their Derivatives Back Propagation
- Deep Learning
- Gradient Recursion for Deep Network Back Propagation
- L^1 Regularization
- L^2 Regularization
- ElasticNet Regression
- L^p Regularization (Generalized Maximum a Posteriori technique)
- Modern Regularization Techniques for Deep Learning
    - Dropout
    - Noise Injection
    - Batch Normalization
- Training with Big Data
  Full vs. Batch vs. Stochastic Gradient Descent
- Techniques for Training Acceleration
    - Variable/Decaying Learning Rates
    - Adaptive Learning Rates
        - AdaGrad
        - RMSProp
        - Adam Optimization
    - Weight Initialization
- Vanishing and Exploding Gradients Optimal Weight Initializations
- Modern Deep Learning Frameworks
    - Theano
    - TensorFlow

- Keras, PyTorch and Sci-Kit Learn

Week 4: Convolutional Neural Networks:

- Signal Processing Theory
- Convolution
  - Echo
  - Gaussian Blurring
  - Edge Detection
- Convolutional Neurons
- Translational Invariance
- Architecture of a Convolutional Unit

Week 5: Unsupervised Machine Learning: Cluster Analysis

- K-Means Clustering
- Soft K-Means Clustering
- The Weighted Distance (or Distortion) Objective Function
- Weaknesses of K-Means Clustering
- Cases where K-Means Clustering can fail
- Gaussian Mixture Models
- Hidden Effects and Latent Random Variables
- Expectation Maximization
- Kernel Density Estimation

Week 6: Supervised Machine Learning: Support Vector Machines

- Support Vectors
- Kernels and Nonlinearity
- "Large Margin" Intuition
- The Mathematics of Large Margin Classification
- The Large Margin Optimization Objective Function
- Support Vector Regression Models

Week 7: Machine Learning: Ensemble Learning Meta-algorithms

- A deeper look at Bias vs. Variance Trade-off
- Decision Trees
- Bootstrapping

- Bootstrap Aggregating (Bagging)
- Bagging Regression Trees
- Bagging Classification Trees
- Stacking
- The Random Forest Algorithm
- Regression Forests
- Classification Forests
- Random Forest vs. Bagged Trees Boosting
- The Ada Boost Algorithm
- Additive Modeling
- Boosting vs. Stacking
- Connections to Deep Learning

Week 8: Distributed Computing and Apache Spark

- Resilient Distributed Datasets (RDDs)
- Spark Data Frames
- RDDs vs Data Frames
- Grouping and Aggregating Operations
- Missing Data
- Dates and Timestamps
- Cloud Computing and Third Party Cloud Resourcese: AWS, Google Cloud, MS Azure
- PySpark API
- SparkR API