Python Machine Learning

GITA315

Professor Emeritus Myungwhan Choi
Department of Computer Science & Engineering
Sogang University

What is this course about?

- Theoretical concepts behind popular machine learning algorithms will be covered
- How learning algorithms are implemented in Python will be shown.
- Insight into how machine learning actually works can be obtained.
- First half portion of the full Python Machine learning course including deep learning
- Will also learn how Numpy, scikit-learn, pandas, TensorFlow2 are used

Course Overview:

big picture

- Giving computers the ability to learn from data (Ch. 1)
- Training simple machine learning algorithms for classification (Ch. 2)
- Tour of machine learning classifiers using scikit-learn (Ch. 3)
- Building good training datasets – Data Preprocessing (Ch. 4)
- Compressing data via dimensionality reduction (Ch. 5)
- Learning best practices for model evaluation and hyperparameter tuning (Ch. 6)
- Combining different models for ensemble learning (Ch. 7)
- Applying machine learning to sentiment analysis (Ch. 8)
- Embedding a machine learning model into a Web application (Ch. 9)
- Predicting continuous target variables with regression analysis (Ch. 10), if time permits
- Working with unlabeled data – clustering analysis (Ch. 11), if time permits

Course Overview:

- Giving computers the ability to learn from data (Ch. 1)
  - Building intelligent machines to transform data into knowledge
  - Using Python for machine learning
- Training simple machine learning algorithms for classification (Ch. 2)
  - Artificial neurons
  - Implementing a perceptron learning algorithm in Python
  - Adaptive linear neurons and the convergence of learning
- Tour of machine learning classifiers using scikit-learn (Ch. 3)
  - Modeling class probabilities via logistic regression
  - Maximum margin classification with support vector machines
  - Solving nonlinear problems using a kernel SVM
  - Decision tree learning
  - K-nearest neighbors

Course Overview:

- Building good training datasets – Data Preprocessing (Ch. 4)
  - Dealing with missing data
  - Handling categorical data
  - Selecting meaningful features
  - Assessing features importance with random forests
- Compressing data via dimensionality reduction (Ch. 5)
  - Unsupervised dimensionality reduction via principal component analysis
  - Supervised data compression via linear discriminant analysis
  - Using kernel principal component analysis for nonlinear mappings
- Learning best practices for model evaluation and hyperparameter tuning (Ch. 6)
  - K-fold cross-validation
  - Debugging algorithms with learning and validation curves
  - Fine-tuning machine learning models via grid search

Course Overview:

- Combining different models for ensemble learning (Ch. 7)
  - Learning with ensembles
  - Combining classifiers via majority vote
  - Bagging
  - AdaBoost
- Applying machine learning to sentiment analysis (Ch. 8)
  - Preparing the IMDb movie review data for text processing
  - Introducing the bag-of-words model
  - Training a logistic regression model for document classification
- Embedding a machine learning model into a Web application (Ch. 9)
  - saving the current state of a trained machine learning model
  - Developing a web application using the Flask web framework

Course Overview:

- Predicting continuous target variables with regression analysis (Ch. 10), if time permits
  - Simple and Multiple linear regression
  - Implementing an ordinary least squares linear regression model
  - Dealing with nonlinear relationships using random forests
- Working with unlabeled data – clustering analysis (Ch. 11), if time permits
  - Grouping objects by similarity using k-means
  - Organizing clusters as a hierarchical tree
  - Locating regions of high density via DBSCAN

- Wide applications areas of machine learning technologies
- Many different ML algorithms, technologies, and associated theories
- implementation and training experience

Course Information

- Who is this course for?
  - MS students
- no prerequisite but helpful:
  - Linear algebra
  - Probability and Statistics
  - Random Variables
  - Optimization Theory
  - Python

- Course materials:
  - Class notes
  - text book

    S. Raschka and V. Mirjalili, Python Machine Learning: Machine learning and Deep learning with Python, scikit-learn, and TensorFlow2, 3rd Ed., Packt, 2019.
  - references
    - M. Mohri et al., Foundations of Machine Learning, The MIT Press, 2012.
    - Tom M. Mitchell, Machine Learning, McGraw-Hill, 1977.

Course Information (more)

- Class WWW site:

  sogang university, cyber campus

- *everything* is posted on this site!
  - syllabus
  - class notes (pdf)
  - exercises and/or assignments if any
- Workload:

Coursework	approx amount	approx %
reading assignments		-
midterm exam		30%
final exam		40%
Lab/assignments and others		30%

- in-class style: interaction, questions *(please!)*