RONDHUIT

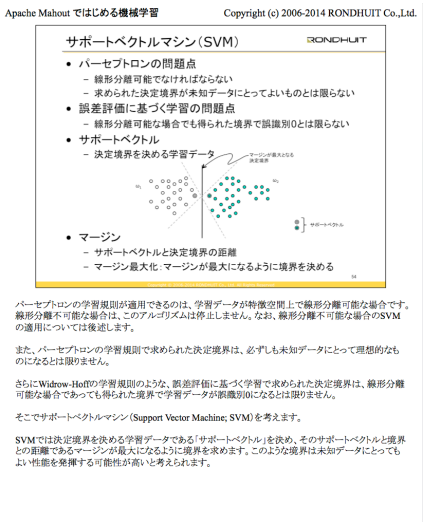
Apache Mahout & Spark ではじめる機械学習

ビッグデータを解析して新しい知見を得ようという取り組みが多くの企業で行われています。そこで不可欠な知識として機械学習が注目されています。

ロンウイット開発によるトレーニングコース **Apache Mahout & Spark ではじめる機械学習**は、ハンズオン（実習）中心のトレーニングコースです。基本的な機械学習に関する知識が体系立てて整理されており、適宜 Mahout, Spark を使う構成になっています。小単元ごとに適切に用意された演習問題を解きながら進められるので、確実に理解が深まり実業務へ応用するための足がかりをつかむことができます。

トレーニングコースの特長

- 限られた時間内で効率よく学習できるよう、テキストには図表がふんだんに使われています。また、毎ページの詳しいノートは帰社後の独習に大いに役立ちます。


Apache Mahout ではじめる機械学習 Copyright (c) 2006-2014 RONDHUIT Co.,Ltd.

サポートベクトルマシン（SVM） RONDHUIT

- パーセプトロンの問題点
  - 線形分離可能でなければならない
  - 求められた決定境界が未知データにとってよいものとは限らない
- 誤差評価に基づく学習の問題点
  - 線形分離可能な場合でも得られた境界で誤識別0とは限らない
- サポートベクトル
  - 決定境界を決める学習データ
- マージン
  - サポートベクトルと決定境界の距離
  - マージン最大化：マージンが最大になるように境界を決める

パーセプトロンの学習規則が適用できるのは、学習データが特徴空間上で線形分離可能な場合です。線形分離不可能な場合は、このアルゴリズムは停止しません。なお、線形分離不可能な場合のSVMの適用については後述します。

また、パーセプトロンの学習規則で求められた決定境界は、必ずしも未知データにとって理想的なものになるとは限りません。

さらにWidrow-Hoffの学習規則のような、誤差評価に基づく学習で求められた決定境界は、線形分離可能な場合であっても得られた境界で学習データが誤識別0になるとは限りません。

そこでサポートベクトルマシン（Support Vector Machine; SVM）を考えます。

SVMでは決定境界を決める学習データである「サポートベクトル」を決め、そのサポートベクトルと境界との距離であるマージンが最大になるように境界を求めます。このような境界は未知データにとってもよい性能を発揮する可能性が高いと考えられます。

- 各単元ごとに用意されたやさしい演習問題を講師と一緒に解くことで、背景理論の考え方や実践的知識が確実に身につきます。

演習問題の例

周長が一定の長方形のうち、面積が最大になるのは正方形であることを、次の 2 つの方法で示してみましょう。

- 連立方程式によるシンプルな変数消去による解法
- ラグランジュの未定乗数法による解法

こんな方におすすめ

- 今注目の機械学習を体系立てて学び、今後の開発業務に役立てたい方。
- ビッグデータを保有する企業の情報処理担当者であり、インテグレーターにビッグデータを活用する開発案件を発注するのに必要な最低限の知識を身につけておきたい方。
- 機械学習系の書籍を購入して勉強したいが、数式が出てきてなかなか読み進められない方。
- Mahout イン・アクションを購入して Mahout を使っているが、いまいち使えている気がしない方。

トレーニングコースの概要

日数および時間	全 2 日間、各日 10:00〜17:00
価格	198,000 円（税別）
トレーニングセンター	ロンウイット清澄白河分室
最少開講人数	2 名

105-0003 東京都港区西新橋 1-18-6 TEL:03-5288-5927 FAX:03-5288-5928
クロスオフィス内幸町 11 階

コース内容

【1 日目】

初日は「機械学習とは何か」から始まり、パターン認識、教師あり学習のいろいろな分類アルゴリズム、そして最後に手書き文字認識プログラムを作成します。受講者全員参加で手書き文字データを作成します。Mahout の分類器は果たしてどのくらい手書き文字を認識するでしょうか？お楽しみに！

- 機械学習と Apache Mahout / Spark MLlib
  - 機械学習とは？モデルとは？
  - Apache Mahout / Apache Spark のインストール、他
- パターン認識
  - パターン認識とは？
  - 特徴ベクトル、距離測度、他
- 分類
  - 最近傍決定則、k-NN 法
  - パーセプトロン、ニューラルネットワーク、サポートベクトルマシン、決定木、単純ベイズ
- 手書き文字認識プログラムを作ろう！
  - 手書き文字認識プログラムの構成
  - 手書き文字データ（訓練データ）の作成、他

【2 日目】

2 日目は Mahout が提供する分類以外の機能であるレコメンデーション、クラスタリングから始まり、特徴ベクトルの次元削減を目的とした主成分分析、機械学習の評価に関する話、そして最後に自然言語処理における機械学習について Mahout / Spark がどのように使えるか、演習を通じて学んでいきます。

- レコメンデーション
  - レコメンデーションとは？
  - 情報検索とレコメンデーション
  - レコメンデーションアーキテクチャの種類
  - ユーザプロファイルとその収集
  - 評価値予測、他
- ページランク
  - ランキングの重要性
  - 情報検索システムの評価尺度の理論と実際
  - ベクトル空間モデル、Apache Lucene のスコア計算
  - ページランク、他
- クラスタリング
  - クラスタリング手法、k 平均法、最近隣法
  - クラスタリング結果の評価と分析、他
- 主成分分析
  - 主成分分析とは？
  - 平均と分散、共分散行列、固有値、固有ベクトル
  - Mahout による主成分分析の実際
- 機械学習の評価
  - 評価指標、特徴の評価、ベイズ誤り確率、他
- 自然言語処理における機械学習
  - 隠れマルコフモデル、ビタビアルゴリズム、他

前提知識

- 演習では Ubuntu マシンを使用しますので、vi や Emacs などのエディタが使えたり、Linux コマンドを知っているとスムーズに受講できます。

持ち物

- ssh が使えるノート PC をご用意ください。ノート PC がご用意できない場合はお貸し出しいたします。
- 手計算による演習問題があるため、鉛筆／シャープペンシル、消しゴムをご用意ください。

お客様のご参加をお待ち申し上げております。

（2014.10.16 版）