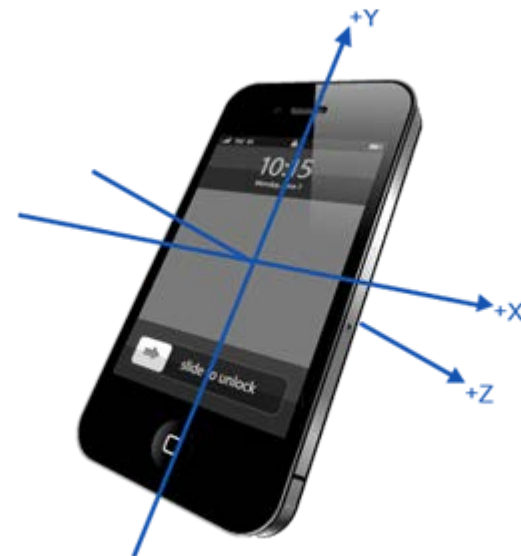


第二回レポート課題（クラス分類）

- Human Activity Recognition Using Smartphones Sensor Data

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012.

- スマートフォンの慣性センサデータ等からユーザの行動を推定
- 30人の被験者データ
- 特徴量（説明変数）：561種類
- 行動カテゴリ（目的変数）：6種類
 - 1. WALKING
 - 2. WALKING UPSTAIRS
 - 3. WALKING DOWNSTAIRS
 - 4. SITTING
 - 5. STANDING
 - 6. LAYING



<http://classmethod.s3.amazonaws.com/wp-content/uploads/2012/06/gyrocompass.png>

配布物

- 学習データ
 - X_train.csv : 特徴ファイル
 - y_train.csv : カテゴリID
 - subject_train.csv : 被験者ID
- テストデータ
 - X_test.csv : 特徴ファイル
- サンプルプログラム
 - sample1_knn.py k最近傍法
 - sample2_svm.py サポートベクターマシン
- その他
 - activity_labels.txt カテゴリ名
 - features_info.txt, features.txt 特徴量の説明
 - authors.pdf 元論文

コンペティション形式

- 被験者IDでデータを学習用・テスト用にランダムに分割
 - 学習データ：22人分 （全7600サンプル）
 - テストデータ：8人分 （全2699サンプル）※元の論文とは分け方が違うので注意
- 評価指標
 - 識別正解率（Accuracy） = 正解したサンプル数 / 全サンプル数
- ランキング
 - スコアリングサーバ: <http://www.nlab.ci.i.u-tokyo.ac.jp/~nakayama/ds13/report2/index.php>
 - テストサンプルの識別結果を一行ずつ記載したテキストファイルを提出
 - ユーザ名を忘れずに（区別できればなんでもよい。）
 - 提出した結果は上書きされる。最後のものだけ保存されるので注意。
 - 現在は、提出されたテストデータのうち、所定の被験者4人分のサンプルでスコアリングしている
 - 最終的なスコアは、締め切り後に残りの4人分のサンプルで算出

課題詳細

- レポート内容
 - Human Activity Recognition の問題に取り組み、以下の点を中心にA4用紙1~2枚程度にまとめよ。講義で扱っていない技術を用いても構わない。
 - 1. 識別性能を向上させるための自分なりの工夫点と結果、考察（必須）
 - コードを載せる必要はない（もちろん、実装がポイントの場合は載せてかまわない）
 - 2. その他、自由にデータを分析した結果（+α）
 - 学習データ数と性能、正則化パラメータの関係
 - 各特徴の寄与の分析
 - カテゴリごとの識別率 …など
 - 3. 講義の感想、要望など

課題詳細

- 実装が難しい場合の課題
 - 配布している元論文あるいは、データマイニング・機械学習に関する論文を一つ自由に選び、内容をまとめよ。また、それが自分の研究にどのように応用できそうか考察せよ

課題詳細

- 評価の方針
 - アイデアや試行錯誤の過程を重視
 - スコアが悪くても、しっかり考察してくればOK
(もちろん良くなればプラスに評価します)
 - 面白い分析を期待します
- 提出先
 - 以下のアドレスへメールで提出すること（質問もこちらへ）
 - ds2013@nlab.ci.i.u-tokyo.ac.jp
 - 件名は「データサイエンスレポート課題2」
 - 氏名、学籍番号、所属を忘れずに

締め切り

- レポート提出締め切り **2月15日（土）**
- コンペについて
 - 2月4日（火） 18:00に一度テストサンプルで中間評価
 - ちょっと変則的ですが、最終日（2月5日）に一度みんなで結果を見たいので
 - 2月11日（火） 18:00にもう一度ランキング（最終）
 - いいレポートのアイデア等は、本人に了解を取りつつ、ホームページに掲載したいと思います。