

レポート課題 2

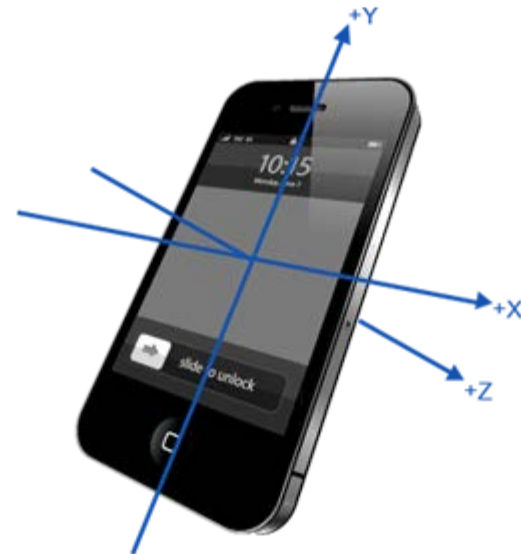
- 内容：下記のうち一つ（選択制）
 - 1. Kaggle のコンペティションをどれか一つ選び、参加せよ。
<https://www.kaggle.com/>
 - 2. 講義で提供するHuman Activity Recognition タスク（後述）
- 以下の点を中心にA4用紙2~4枚程度にまとめよ。
講義で扱っていない技術も自由に用いて構わない。
 - 1. タスクの詳細（特にKaggle参加の場合）
 - 2. 試した方法、性能を向上させるためのアイデア・工夫点など
 - 3. 結果と考察
 - 4. 講義の感想、要望など
- ✂切
 - 2/13（火） 23:59

コンペ課題

- 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：15人分 （全5080データ）
 - テストデータ：15人分 （全5219データ）※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） = 正解したデータ数 / 全データ数
- ランキング
 - スコアリングサーバ: <http://www.nlab.ci.i.u-tokyo.ac.jp/~nakayama/ds17/report2/index.php>
 - テストデータ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: 学籍番号を一部に入れること (例:48123456method1)
 - 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(validation)は、提出されたテストデータのうち、所定の被験者5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、締め切り後に残りを加えたの15人分のサンプルで算出

その他

- 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)
- 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）
 - ds2017@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
- レポート提出締め切り 2月13日（火） 23:59
- コンペについて
 - 締め切り終了後にテストデータで最終評価
 - サーバへ提出するユーザ名に学籍番号(ID)を入れてください！