

ステップ2：学習させる

ご購入はこちら

佐藤 聖

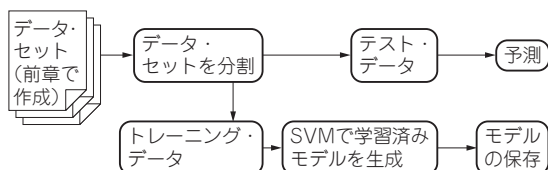


図1 人工知能の重要工程…データ・セットから学習済みモデルを作る

第1章では撮影対象物の画像を使ってデータ・セットを作りました。本章では、このデータ・セットを利用して学習を行い、「学習済みモデル」を作っていきます。

モデルの構築にはサポート・ベクタ・マシン(以降、SVM)を使います。SVMが開発された当初は、線形分離器(各データ点間のマージンが最大になるように直線を引いて分類)にしか適用できませんでしたが、改良が重ねられて線形分離が不可能な問題(非線形分離問題)にも適用できるようになりました。そのため汎用分離器として幅広い用途に応用が可能です。

ここで言うことは、図1(Appendix1 図1の一部を再掲載)にあるように、データ・セットをSVMを用いて学習し、学習済みモデルを構築した後に、ファイルに保存します。

05 SVMのトレーニングと学習モデル作成. ipynb(リスト1)では、データ・セットからトレーニング・データとテスト・データに分割して学習モデルを構築します(図2)。

用意したデータから「トレーニング・データ」と「テスト・データ」を生成

● ライブラリの読み込み

最初にライブラリを読み込みます(リスト1のIn [1])。体験サンプルCのSVMとscikit-learnのcross_validationライブラリを利用しているので、Out [1]に警告メッセージが表示されますが動作には問題ありません。

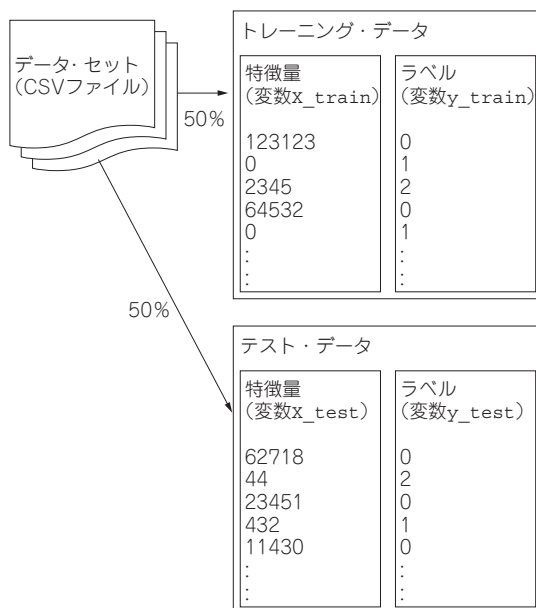


図2 データ・セットをトレーニング用とテスト用に分ける

● CSVファイルを読み込み

CSVファイル(pic_data.txt)からデータを読み込みます(In [2])。読み込んだデータから特徴量(hist列)とラベル(label列)を、トレーニング・データとテスト・データに加工します。

● 透過画像を抽出して列名を付け替え

CSVファイルのデータはまず透過画像と表面画像が同じ列に混在しているため、透過画像と表面画像のhist列を同じシーケンス番号(seq_num)と角度(angle列)の組み合わせごとに1行になるように加工します(In [3])。最初に透過画像側のデータだけを抽出してデータ・フレームback_tmpに格納されるようにします。

データ・フレームdfを操作してside列が0の行だけ抽出してデータ・フレームbackを作成し、さらにside列なしのデータ・フレームにするためloc