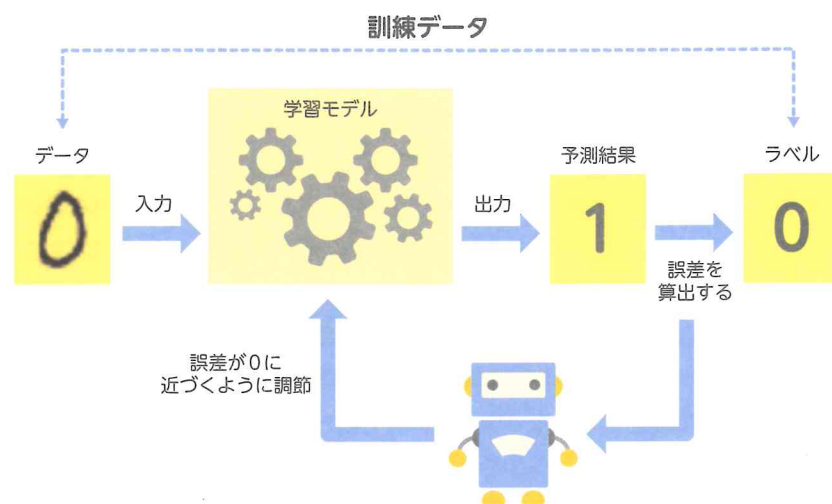


教師あり学習は誤差を小さくする

冒頭で記したように、教師あり学習の最終目標は、ラベルのないテストデータを正解させることです。そのためには、まずラベルのある訓練データをモデルが正解するようにしなくてはなりません。言い換えれば、モデルが出力した予測とラベルとの誤差を少なくしなくてはなりません。実際の学習では、これらの誤差を0に近づけることで正解できるデータの数を増やしていきます。なお、分類では交差エントロピー誤差、回帰では平均二乗誤差と呼ばれる誤差がよく使われます。

教師あり学習は誤差を小さくする



まとめ

- 教師あり学習とは、教師データをモデルに学習させる方法のこと
- 教師あり学習の最終目標は、テストデータを正解すること
- 教師あり学習は分類と回帰に分けられる

COLUMN

エンジニアのスキルアップに欠かせないKaggleとは？

Kaggle (<https://www.kaggle.com/>) とは、世界中のデータサイエンティストや AI エンジニアが集まる約 40 万人のコミュニティです。とりわけ注目すべきは、企業や政府の課題に対して参加者が最適なモデルを提出し、最も優れたモデルに賞金が支払われる「コンペ」ことコンペティションです。コンペティションでは教師あり学習が採用されており、参加すればデータセットをダウンロードすることができます。コンペ1つの開催期間は3~6ヶ月で、リーダーボードにリアルタイムで順位が更新されるため、参加者のモチベーションに直結するしくみとなっています。

そのほか Kaggle では、ブラウザ上でコードを動かすことができる「kernel」という機能でコンペに対するデータ解析の結果を共有できます。また「discussion」という議論の場も用意されています。そのため、コンペにモデルを提出しなくても十分な知見が得られるでしょう。ただし英語のサイトなので、読み進めるのに少し時間がかかるかもしれません。

同様のサイトとして SIGNATE という日本のサイトもありますが、こちらは Kaggle ほど活発ではなく、母国語での熱心な議論にはもう少し時間がかかりそうです。

Competitions

Documentation

InClass

General

2019-04-04


Sort by

Grouped

All Categories

Search competitions

19 Active Competitions




Two Sigma: Using News to Predict Stock Movements

Use news analytics to predict stock price performance

Featured | Kaggle Competition | 4 months to go | News agencies, some banks, finance firms.

\$100,000

2,927 teams




Santander Customer Transaction Prediction

Can you identify who will make a transaction?

Featured | Kaggle Competition | 1 month to go | Banking, financial data, data science, finance.

\$65,000

4,901 teams



LANL Earthquake Prediction

Can you predict upcoming laboratory earthquakes?

Research | Kaggle Competition | 1 month to go | Big data, computing, physics, earth sciences.

\$50,000

1,641 teams