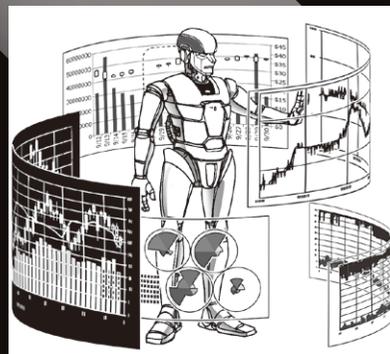


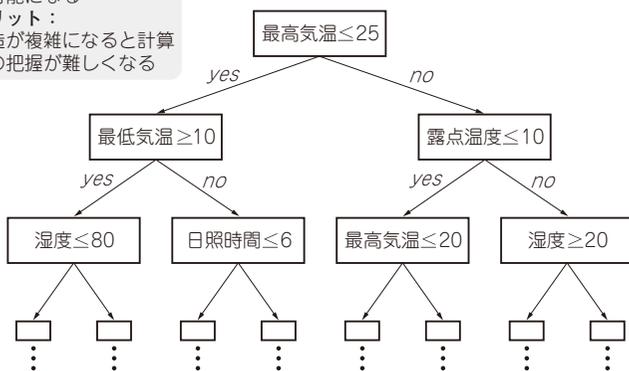
脱! ブラックボックス 説明可能なAI

第3回 どの特徴量が貢献しているかを表す Feature Importance

切通 恵介

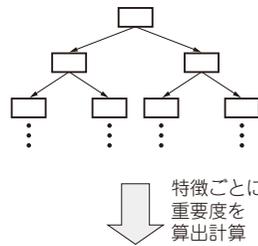


メリット:
計算過程を全て把握することができるため、正確な説明が可能になる
デメリット:
木構造が複雑になると計算過程の把握が難しくなる



(a) 一般的に使われる手法…木構造の可視化

メリット:
木構造が複雑になっても特徴の重要性を俯瞰的に理解できる
デメリット:
統計量に近く、細かな計算過程を知ることができない



特徴	重要度
最高気温	0.5
最低気温	0.2
湿度	0.1
露点温度	0.05
日照時間	0.02
⋮	⋮

(b) 今回紹介する手法…Feature Importance

図1 決定木の説明手法

ディープ・ラーニングや機械学習などのAIモデルは、複雑性が高くAIモデルの中で何が起きているのかは、多くの場合、ブラックボックスとして扱われます。そんな中で説明可能なAI (eXplainable AI: XAI) という概念が生まれました。説明可能とは、AIモデルの説明が得られる、すなわち、AIモデルが出した結果の根拠を示せるということです。結果の根拠とは、モデルへの入力特徴の重要度を数値化したものや、計算過程を可視化したものです。本連載では、AIモデルが出した結果の根拠を計算するアルゴリズムとその実装までを解説します。

● 今回のテーマ

今回は、決定木回帰モデルを元に計算される説明手法の1つであるFeature Importanceを取り上げます。使用例として、決定木モデルの拡張として最も利用されているモデルの勾配ブースティング木で、Feature Importanceがどのように計算されるかを解説します。

● 決定木とは

決定木は、データの特徴ごとにしきい値で分割することを繰り返すことで木構造を作成し、その最後の葉の部分で得られる結果を用いて予測する機械学習モデルの1つです。決定木を使った予測の説明、すなわち計算過程やどの特徴が重要視されているかを可視化するに当たって一般的に使われる手法は、図1(a)のように、決定木がどの特徴、どのしきい値で分割しているかの構造を可視化するものです。

決定木の欠点…木構造が深くなると解釈が難しくなる

決定木は計算過程がはっきりしており、厳密にモデルの振る舞いを判断できる反面、決定木の深さが深くなるにつれて木構造の複雑さが増すため、説明の解釈が難しくなるという欠点があります。例えば、木の深さが100層になったとき、分割の数は 2^{100} という途方もない数になります。この膨大な木構造をつぶさに見ていって、どの分割が重要であるかの全容を把握するのは至難の業です。