



Visual Studio Code & LightGBM 編

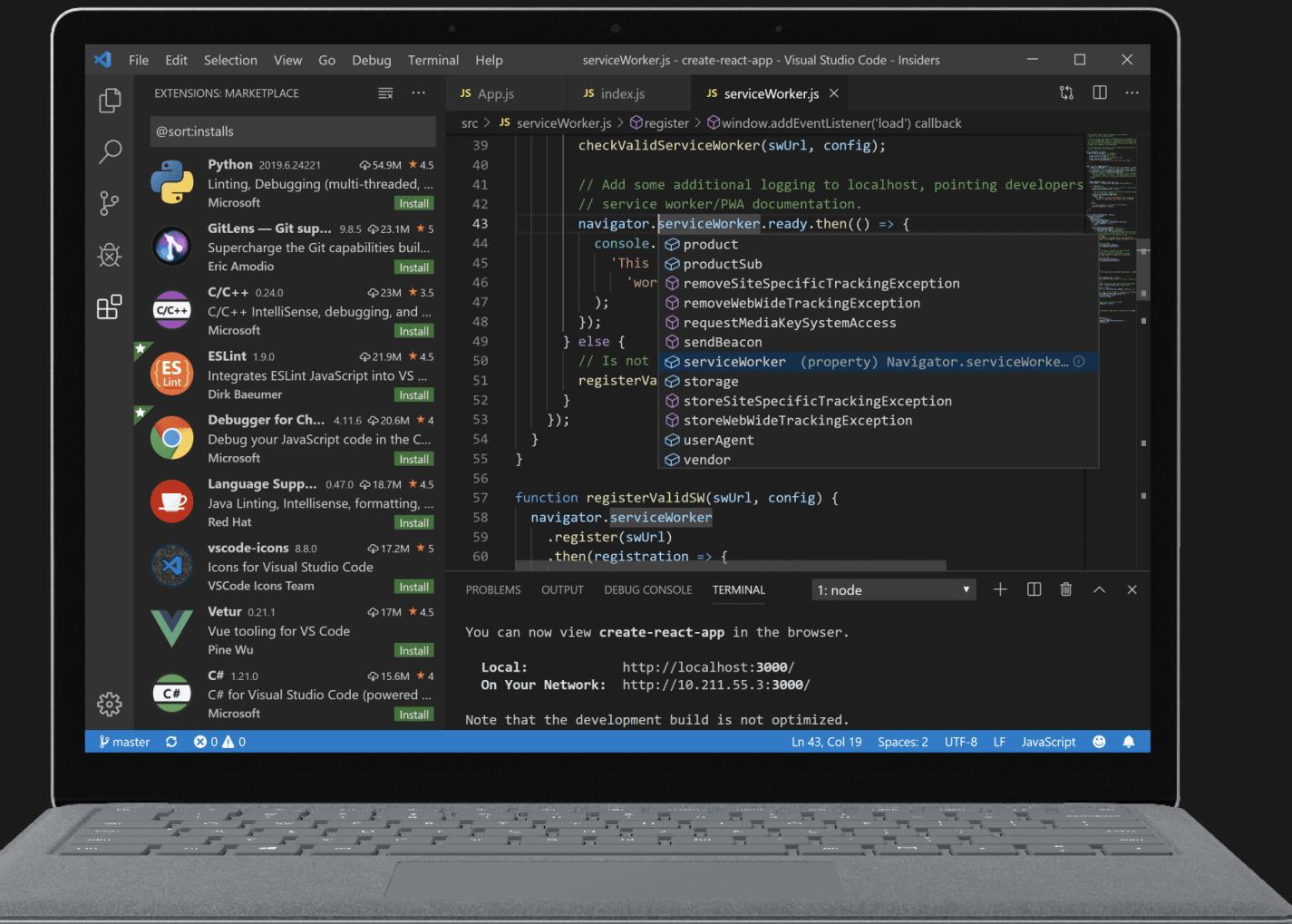
Azure Machine Learning ハンズオン

女部田啓太、Cloud Solution Architect -Data & AI

アジェンダ

- 環境セットアップ
 - Visual Studio Code
 - Azure Machine Learning
- LightGBM ハンズオン
 - ボストン住宅価格予測
- 参考情報

環境セットアップ



Visual Studio Code

Edit and debug across
macOS, Linux, and Windows

Strong ecosystem of over 16k
extensions, themes, and tools

Work in containers, WSL, or over
SSH with remote development

Visual Studio Code Extension

必須

- [Python](#)
 - Pylance
 - Jupyter
- [Azure Machine Learning](#)
 - Azure Account

おまけ

- [Docker](#)
- [vscode-pet](#)
- [Power Model](#)

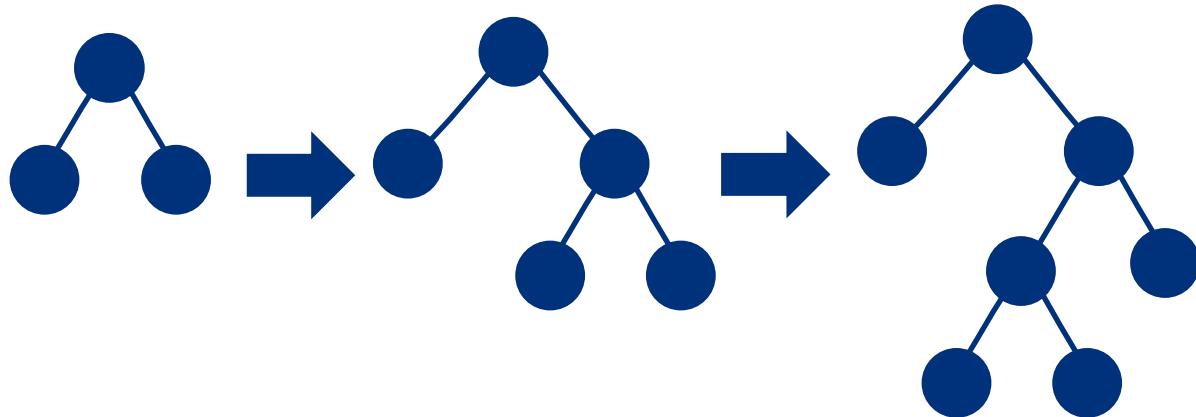
参考ドキュメント

- [Python in Visual Studio Code](#)
- [Get Started Tutorial for Python in Visual Studio Code](#)
- [Visual Studio Code 拡張機能 \(プレビュー\) を設定する - Azure Machine Learning | Microsoft Docs](#)



LightGBM

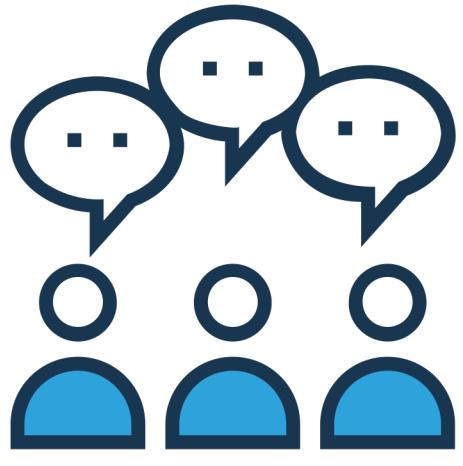
Microsoft が開発した勾配ブースティング手法



- ・複数の決定木ベースの弱学習器
- ・アンサンブル学習の Boosting を利用

Leaf-wise tree growth

Why LightGBM?



kaggle™

実績



Scikit learn
準拠の API



Fast !

勾配ブースティング手法の歴史



2016年に一躍有名に



カテゴリデータに強み

参考資料

- [microsoft/LightGBM \(github.com\)](https://github.com/microsoft/LightGBM)
- [LightGBM's documentation! — LightGBM 3.2.1.99 documentation](https://lightgbm.readthedocs.io/en/latest/index.html)
- [Azure/mmlspark: Microsoft Machine Learning for Apache Spark \(github.com\)](https://github.com/Azure/mmlspark)
- [「初手LightGBM」をする7つの理由 - u++の備忘録 \(hatenablog.com\)](https://hatenablog.com/u++/entry/2018/09/13/113000)

LightGBM ハンズオン

ボストン住宅価格の予測

テーマ：住宅価格予測

列名	説明
crim	人口 1 人当たりの犯罪発生数
zn	25,000 平方フィート以上の住居区画の占める割合
indus	小売業以外の商業が占める面積の割合
chas	チャールズ川によるダミー変数 (1: 川の周辺、0: それ以外)
nox	窒素酸化物 (NOx) の濃度
rm	住居の平均部屋数
age	1940 年より前に建てられた物件の割合
dis	5 つのボストン市の雇用施設からの距離
rad	環状高速道路へのアクセスのしやすさ
tax	10,000 ドルあたりの不動産税率の総計
ptratio	児童に対する教師の比率
a	アフリカ系の比率
lstat	低所得者の比率 (%)
medv	住宅価格の中央値 (単位: 1,000 ドル)

- The Boston Housing Dataset
 - <https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html>
 - 米国ボストン市郊外における地域別の住宅価格のデータセット
 - 506 件、14 属性

ハンズオンの内容 #1

- Goal
 - ・ Azure Machine Learning Compute Instance + Visual Studio Code を用いて、LightGBM モデル学習を行う
- データ
 - ・ ボストン住宅価格
- 手順
 1. Azure Machine Learning Compute Instance を起動
 2. Compute Instance から Visual Studio Code を起動
 3. コードをクローン
 4. Conda 環境の作成
 5. コード実行

ハンズオンの内容 #2

- Goal

- Azure Machine Learning Compute Instance + Visual Studio Code を用いて、LightGBM モデル学習を行う
- Azure ML 実験管理でメトリックを追跡する

- データ

- ボストン住宅価格

- 手順

1. コード実行
2. ハイパーパラメータ、モデル精度のメトリックの確認
登録されたモデルの確認

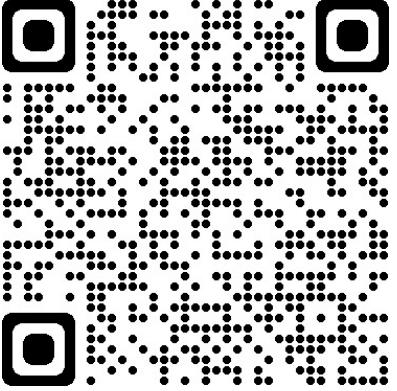
参考情報



Machine Learning Practices & Tips

機械学習プロジェクトを進めるためのガイドブック

[Machine Learning Best Practices \(azure.github.io\)](https://Machine Learning Best Practices (azure.github.io))



Microsoft Machine Learning Collection

Microsoft / LinkedIn / GitHub
機械学習テクノロジーコレクション

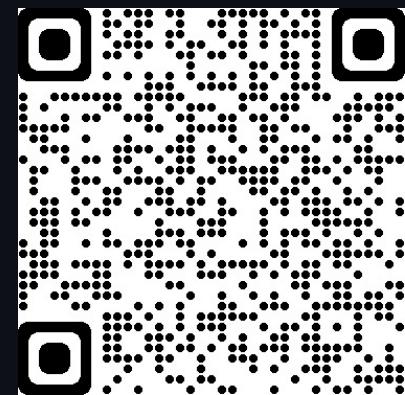
[microsoft/machine-learning-collection:](https://github.com/microsoft/machine-learning-collection)
[machine learning tech collections at Microsoft and subsidiaries.](https://github.com/microsoft/machine-learning-collection)

Machine Learning Collection

Microsoft contributing libraries, tools, recipes, sample codes and workshop contents for machine learning & deep learning.

Table of Contents

- Boosting
- AutoML
- Neural Network
- Graph&Network
- Vision
- Time Series
- NLP
- Interactive Machine Learning
- Recommendation
- Distributed
- Casual Inference
- Responsible AI
- Optimization
- Reinforcement Learning
- Windows
- Datasets
- Debug
- Pipeline
- Platform
- Tagging
- Developer tool
- Sample Code
- Workshop
- Book
- Learning
- Blog, News & Webinar



参考情報

- [Machine Learning Practices & Tips](#)
- [Microsoft Machine Learning Collection](#)
- Microsoft Learn - [Data Scientist 向けページ](#)



Microsoft AI



