



自然言語処理 BERT 編

Azure Machine Learning ハンズオン

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

アジェンダ

- BERT ハンズオン
 - 学習タスクの挙動確認 (MASK, NSP)
 - BERT Fine Tuning
- 参考情報

BERT ハンズオン

学習タスクの挙動確認 (MASK, NSP)
BERT Fine Tuning

自然言語処理 NLP とは

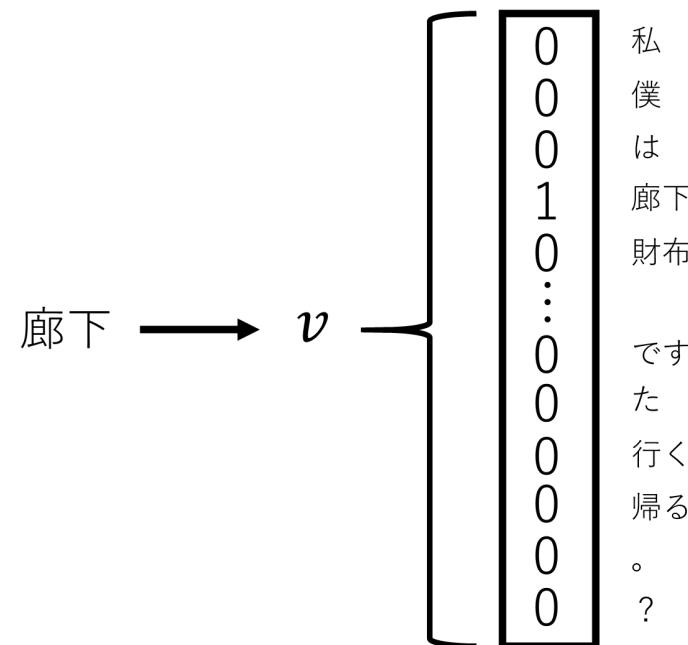
- ・人間が使用する書き言葉や話し言葉等の自然言語をコンピューターで取り扱う一連の技術
- ・1960年頃から今日に至るまで、コンピューターの発展と共に研究が続けられた分野
- ・現在では基礎的な技術の一部に伝統的な統計的手法が残っているものの、自然言語処理全体としては大半が深層学習の領分になりつつある

複合技術

- ・機械翻訳：ある言語で記述された文を別の言語へと変換する
- ・質問応答：質問文に対する回答を生成する
- ・文書分類：文書がどの分類に属するか判別する
- ・感情分析：文書のネガ/ポジ(あるいはもっと細かい分類)を判断する
- ・キャプション生成：画像の説明文を生成する

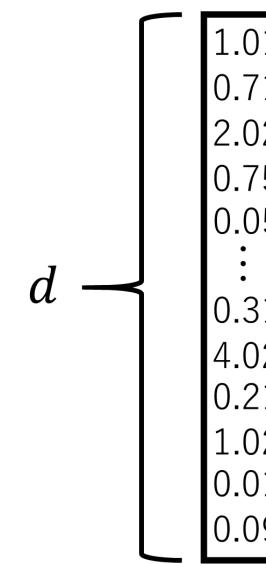
自然言語処理の流れ

単語の one-hot 表現



ベクトルの各次元と単語を対応させ、
0 と 1 で単語の有無を表す形にテキストを変換する

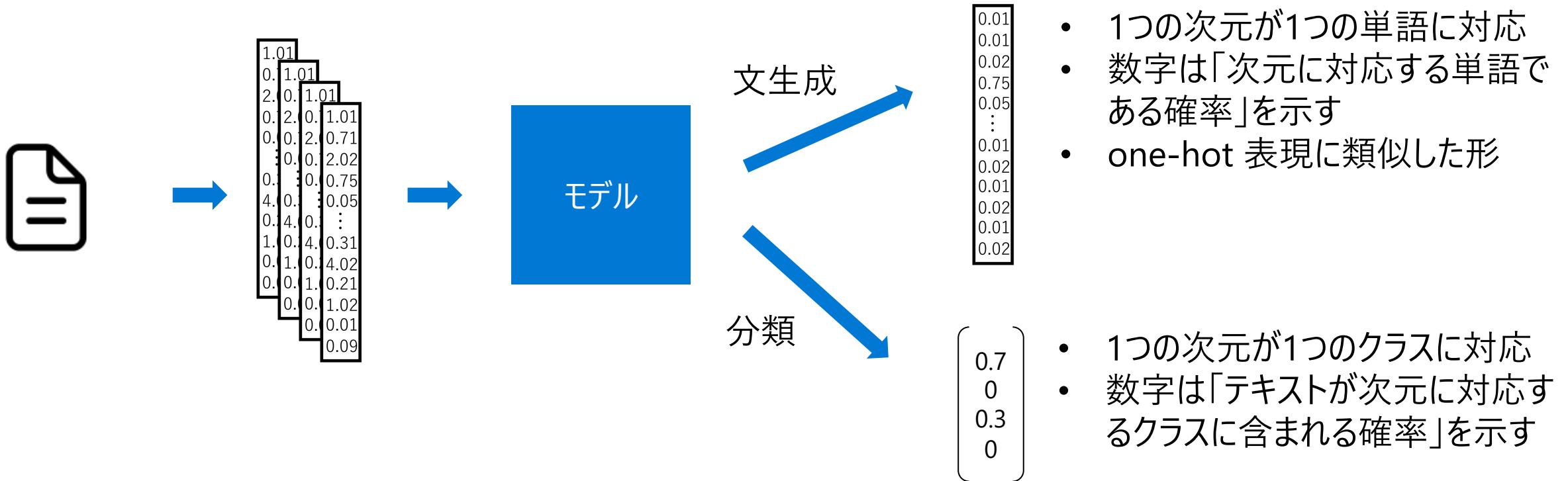
単語の Embedding 表現



Embedding によって one-hot 表現の単語を
よりサイズが小さいベクトルに変換する

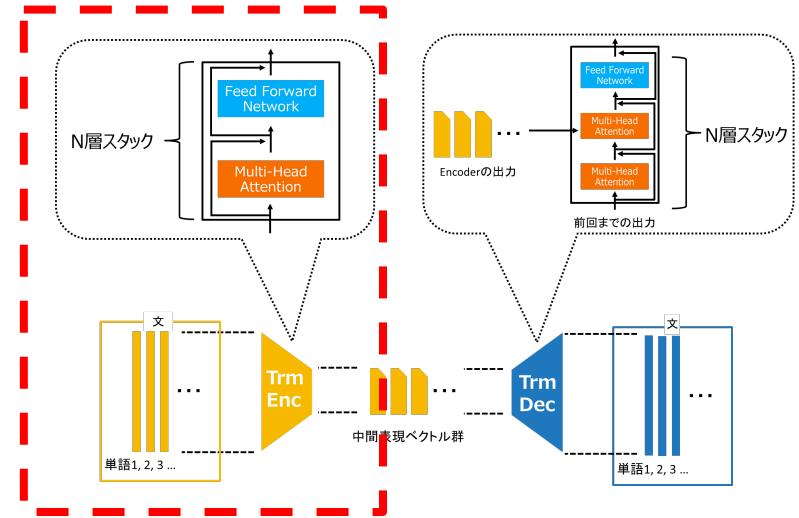
自然言語処理の流れ

分散表現に変換したテキストをモデルに入力し、欲しい出力を得る



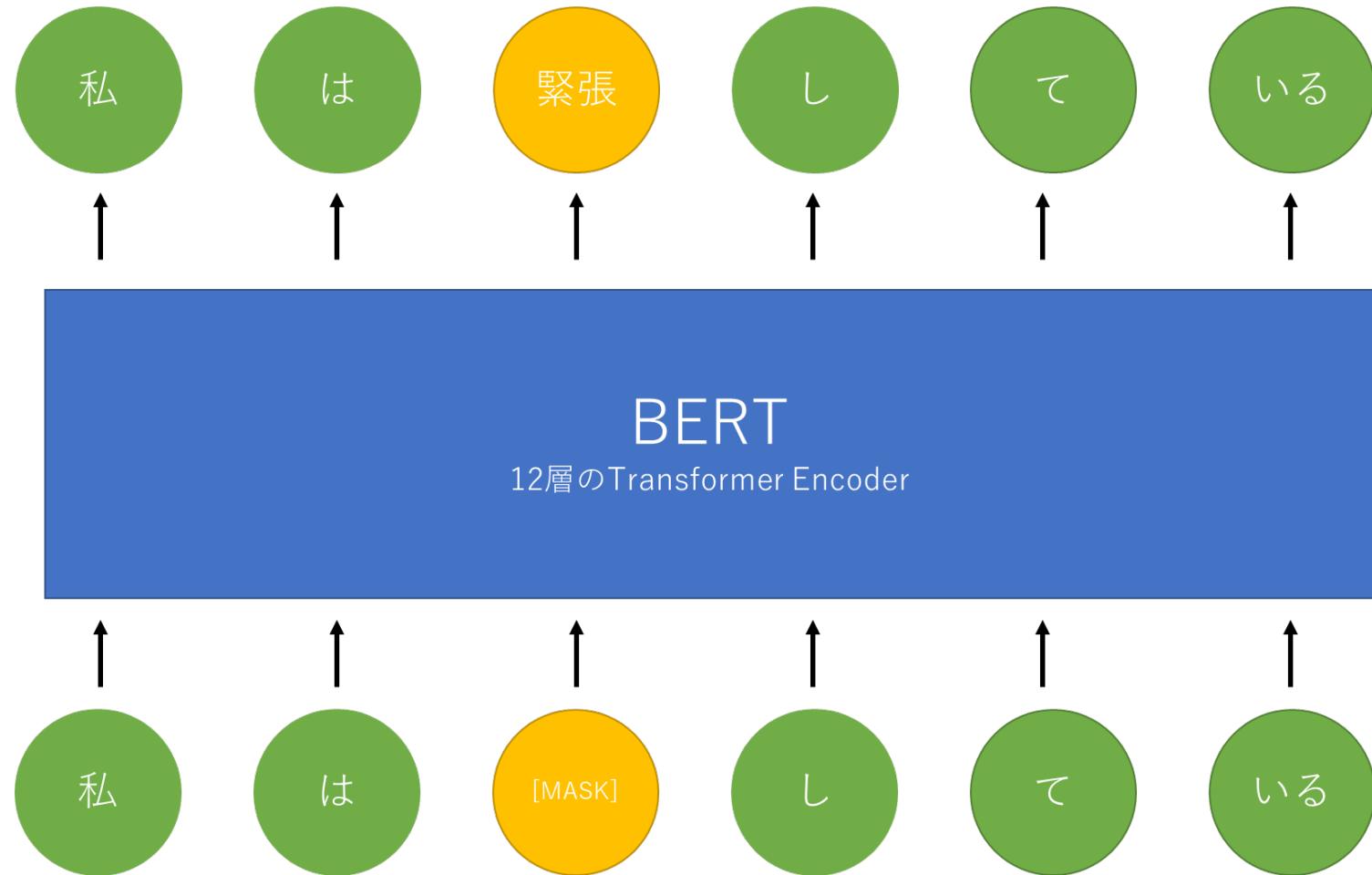
BERT (Devlin et al, 2018)

- ・ 巨大なTransformer Encoder
- ・ 2種類の言語モデル的事前学習
 - ・ マスクされた単語の予測
 - ・ 文の連続判定
- ・ 膨大なデータで事前学習を行った後、少数のデータで目的タスクに転移学習
- ・ NLPの幅広いベンチマーク(分類系タスク)でSoTA



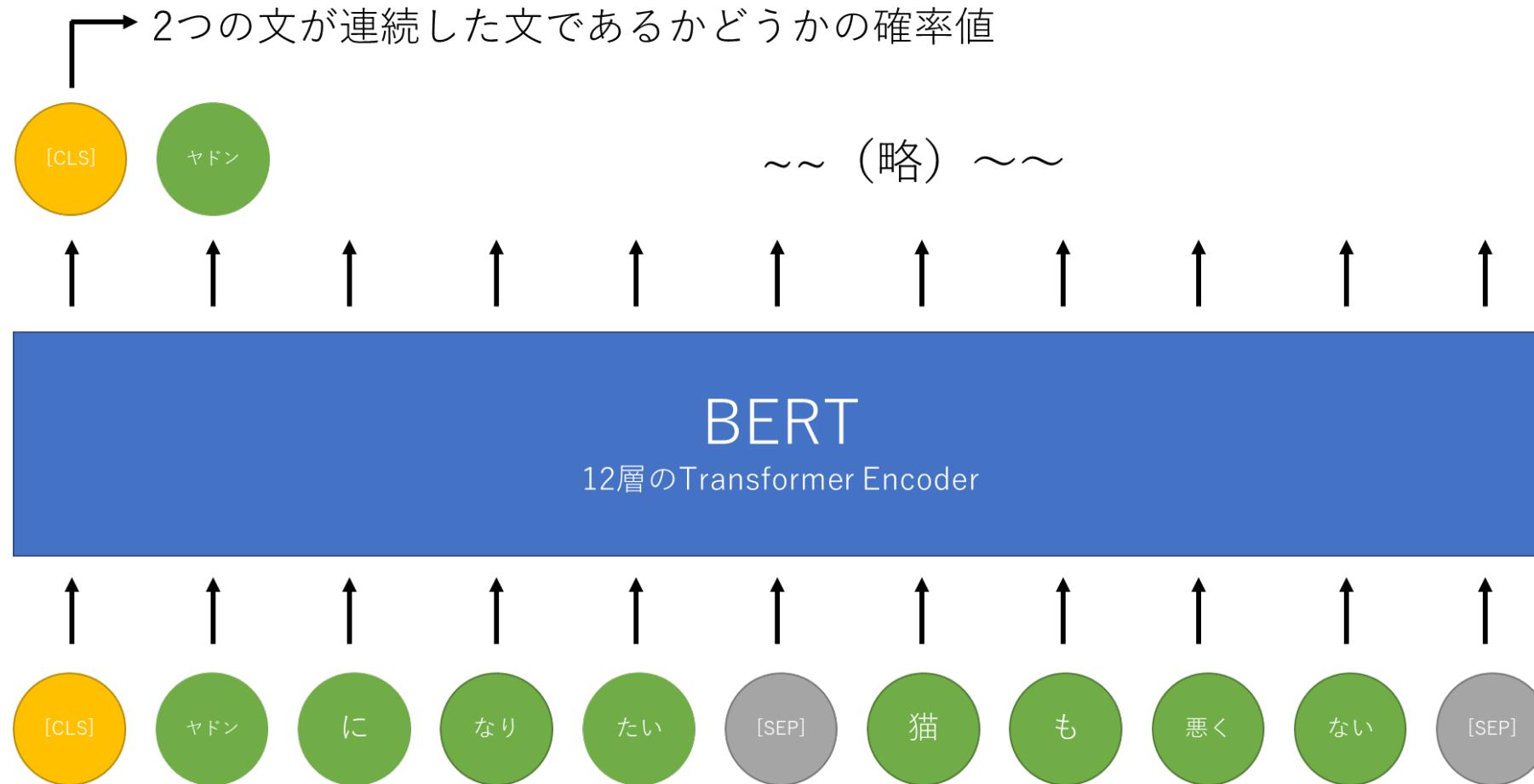
BERTの事前学習タスク①

- マスクされた単語の予測



BERTの事前学習タスク②

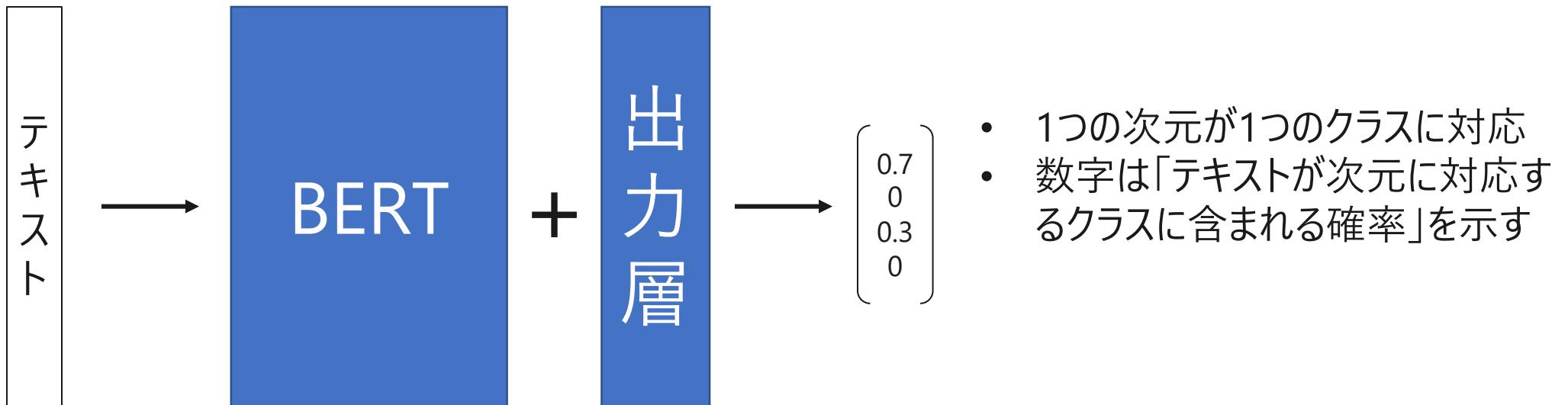
- 文の連続判定



BERT の実タスクへの適用

- Fine-tuning

- 言語モデル学習をした BERT に対して、適用したいタスクのデータで転移学習
- 「BERT 部分は固定し、タスク用のネットワークを追加して学習する方法」と「全て更新する方法」の大きく2通り



BERT (Devlin et al, 2018)

- ・分類タスクにおいて圧倒的な性能を発揮
 - ・11個のタスクで SoTA を叩き出す
 - ・質問応答などいくつかのタスクで人間を上回る
- ・ライブラリや各言語対応モデルが豊富、ノウハウも豊富
- ・分類タスクなら現在も通用する性能で、周辺環境が整っているため非常に応用しやすい
- ・以後の自然言語処理における潮流を「Transformer ベースの巨大モデル + 大量のデータで言語モデル的事前学習」へと完全に塗り替えた

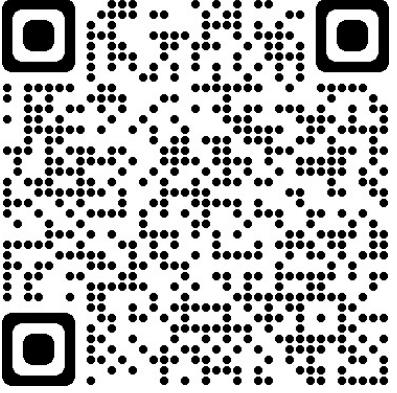
参考情報



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

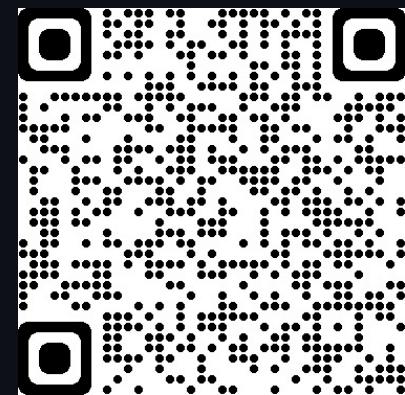
[microsoft/machine-learning-collection:](https://github.com/microsoft/machine-learning-collection)
[machine learning tech collections at Microsoft and subsidiaries.](https://github.com/microsoft/machine-learning-collection)

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](#)
 - [Introduction to Natural Language Processing with PyTorch - Learn | Microsoft Docs](#)



Microsoft AI



