

・異常検出: この種のタスクでは, コンピュータプログラムが一連のイベントあるいはオブジェクトを選別し, 異常であるとしてそれらの一部にフラグを立てる. 異常検知タスクの例としては, クレジットカードの不正検出が挙げられる. 購買習慣をモデル化することで, クレジットカード会社はカードの不正利用を検出できる. 泥棒があなたのクレジットカードやクレジットカードの情報を盗んだ場合, 泥棒の購買はあなた自身の購買タイプとは異なる確率分布に基づいて行われることが度々起こる. クレジットカード会社は, カードが普段とは異なる購入に使用されると即座にアカウントを保留することで, 不正を防ぐことができる. 異常検出手法の調査については, チャンドラを参照.

・合成とサンプリング: この種のタスクでは, 機械学習アルゴリズムはトレーニングデータ内のサンプルに類似した新たなサンプルを生成するよう求められる. 機械学習による合成とサンプリングは, アーティストが手作業で大量のコンテンツを生成するのに費用がかかり, 作業が単調な場合にメディアアプリケーションにおいて役立つ. 例えば, (2013年のルオらによれば) ビデオゲームではアーティストが各ピクセルに手でラベルをつける必要がなく, 大きなオブジェクトや風景のテクスチャを自動的に生成できる. 特定の場合には, 入力に基づいてサンプリングあるいは合成手順が特定の種類の出力を生成することを望むことがある. 例えば, 音声合成タスクでは, 文書を提供し, その文の音声バージョンを含む音声波形を出力するようにプログラムを求める. これは, 一種の構造化された出力タスクだが, 各入力に対して単一の正確な出力が存在しないという条件が付加されており, 出力がより自然で現実的に見えるように出力の大きな変化が明示的に要求される.

・欠損値の補完: この種のタスクでは, 機械学習アルゴリズムに新たな例 $x \in \mathbb{R}^n$ が与えられるが, x のいくつかのエントリ x_i が欠落している. アルゴリズムは欠落しているエントリの予測値を提供する必要がある.

・ノイズ除去: この種のタスクでは, 機械学習アルゴリズムには破損していないサンプル $x \in \mathbb{R}^n$ から未知の破損プロセスによって取得された破損サンプル $\tilde{x} \in \mathbb{R}^n$ が入力として与えられる. 学習者は, 破損したバージョン \tilde{x} から破損していないサンプル x を予測する, もしくはより一般的に条件付き確率分布 $p(x|\tilde{x})$ を予測する必要がある.

・密度推定または確立質量関数の推定: 密度推定問題では, 機械学習アルゴリズムは関数 $p_{\text{model}}: \mathbb{R}^n \rightarrow \mathbb{R}$ を学習するように求められる. ここで, (x が連続の場合) p_{model} は確率密度関数として解釈でき, (x が離散の場合) サンプルが抽出された空間上の確立質量関数として解釈できる. こういったタスクを適切に実行する (すなわち, パフォーマンスの指標 P について説明する際, それが何を意味するかを正確に指定する) には, アルゴリズムは認識したデータの構造を学習する必要がある. サンプルが密集している場所とサンプルが発生する可能性が低い場所を把握する必要がある. 上記のタスクのほとんどは, 学習アルゴリズムが少なくとも暗黙的に確率分布の構造を把握している必要がある. 密度推定により, その分布を明示的に捉えることができる. 原則として, 他のタスクを解決するためにその分布に対して計算を実行できる. 例えば, 密度推定をして確率分布 $p(x)$ を取得した場合, その分布を使用して欠損値を補完するタスクを解くことができる. 値 x_i が欠落しており, x_{-i} で示される他の値が全て与えられているのであれば, その分布は $p(x_i|x_{-i})$ で与えられることがわかる. 多くの場合, $p(x)$ で必要な演算が計算上処理できないため, 実際には, 密度推定によってこれらの関連タスクを全て解決できるとは限らない.

もちろん, 他の多くのタスクや種々のタスクについても可能である. ここで記載した種々のタスクは機械学習で何ができるかの例を提供することだけを目的としており, タスクの厳密な分類を定義するものではない.

1 5.1.2 パフォーマンス指標, P

機械学習アルゴリズムの性能を評価するには, パフォーマンスの定量的な尺度を設計する必要がある. 通常, このパフォーマンス指標 P はシステムによって実行されたタスク T に特化している.

分類, 入力欠損した分類, 転写などのタスクでは, モデルの精度を測定することが度々ある. 精度は, 単にモデルが正しい出力を生成するサンプルの割合のことを表す. また, モデルが誤った出力を生成するサンプルの割合にあたるエラー率を測定することでも同等の情報を得ることができる. エラー率を期待 0-1 の損失と呼ぶことがよくある. 特定のサンプルにおける 0-1 損失は, 正しく分類できていれば 0, そうでなければ 1 である. 密

度推定などのタスクでは、精度、エラー率、その他の種類の 0-1 損失を測定することには意味がない。代わりに、モデルに各サンプルの連続値スコアを与える別のパフォーマンス測定基準を使用する必要がある。最も一般的なアプローチは、モデルがいくつかのサンプルに割り当てる平均対数確率を報告することである。

通常、私たちは機械学習アルゴリズムがこれまで見た事のないデータに対してどの程度うまく機能するかに興味を持つ。これは、現実世界に展開された際に機械学習アルゴリズムがどの程度うまく機能するかを示すからである。したがって、機械学習システムを訓練するテストセットに使用するデータとは別のデータを使用して、これらのパフォーマンス指標を評価する。

パフォーマンス指標の選択は単純で客観的であるかのように思えるが、システムの望ましい動作によく合致するパフォーマンス指標を選択することは難しいことがある。