スライド２　2.1なぜ深層学習か？

　まず，深層学習(ディープラーニング)とは何か？について説明します．近年，AI（人工知能）という言葉が身近になっていますが，このAIという言葉の定義は厳密に決まっておらず，人ぞれぞれ違った定義をしているのが現実です．ただ，一般的な考え方としてAIとは人工知能ということで「人間のような知能を持ったコンピュータ全般」を指すことが多いと思います．つまり，昔，人間が行っていたことをコンピュータに行わせればそれはAIといえるわけです．つまりAIというのは最近登場したものではなく，もう数十年前から存在しているということになります．では近年話題になっているAIというのは何を指すのかというと，それは機械学習・ディープラーニングだと思います．機械学習というのはAIの中の一つの分野です．意味としては「機械(コンピュータ)がデータから自ら学習して，データの背景にあるルールやパターンを発見するもの」です．その機械学習にもいろいろな手法があります．その中で，最もすごいと期待されている手法は深層学習(ディープラーニング)ということです．つまり，図で表すとAI，機械学習，深層学習というのはこのような関係になります．

スライド3　2.1なぜ深層学習か？

　このセミナーでは深層学習の基礎，特に深層学習の計算の中身である「深層ニューラルネットワーク」の基礎について説明するつもりです．

　深層学習とはどういった手法かというと．この本では「動物の神経回路にヒントを得て提唱された (深層) ニューラルネットワーク計算により，大量のデータからその背後に潜む知識を自発的に獲得していく手法」と書かれています．右下に深層ニューラルネットワークの一例を載せました．丸や線を組み合わせた図でニューラルネットワークというのは作られます．丸は「ノード」や「ニューロン」とよばれ，線は「エッジ」や「ウエイト」という風によばれます．物理でいうファインマン図のようなものです．この丸や線1つ1つに数式が割り当てられており，この図で深層学習の複雑な計算を表しています．深層学習とはどういったことを行うかというと，何かデータをニューラルネットワークに計算させると，データからいろいろなことを学習して，我々が知りたいことを結果として出してくれるという流れです．

　深層学習はこういった手法ですが，これの何がすごいのかというと主に2つあります．まず1つ目は汎用的なアルゴリズムを提供してくれるということです．（機械学習と深層学習で比較して説明する）2つ目は高い汎化性能があるということです．この汎化とは何かというと「手持ちのデータだけの中から，すべての状況に通用する本質的な知識を獲得する」ということです．(手書き文字の例を説明する)

　これらの性質を持っているというのが，深層学習が現在期待されている理由です．

スライド4　2.2機械学習とは何か

　深層学習を理解するにはまず機械学習についてある程度知っておく必要があります．ということでここでは機械学習とはなにかについて説明します．機械学習とは「人間がこなすような様々な学習や知的作業を計算機に実行させるためのアプローチの研究，あるいはその手法そのもの」と書かれています．ここで，トム・ミシェル・ミッチェルという方による機械学習の定式化を紹介します．それは「コンピュータプログラムが，ある種のタスク T と評価尺度 P において経験 E から学習するとは，タスク T におけるその性能を評価尺度 P によって評価した際に，経験 E によってそれが改善されている場合である」です．よくわからないと思います．ここで現れたタスクや評価尺度，経験について説明します．(この3つについて説明)．要するに「経験の蓄積によってタスクを解いたときに，評価尺度が向上していくような手法のことを機械学習という」ということです

スライド5　2.2機械学習とは何か

　先ほどタスクというので出てきていくつか紹介しましたが，ここでは代表的なタスクについて説明します．

1つ目はクラス分類です．これはカテゴリ(クラス)に仕分ける作業ということです．例えば電子メールを見て，通常メールかスパムメールかを判断するような作業は2クラス分類ということです．クラス分類では与えられたデータを数値データxとして，クラスをC\_yのようにラベリングし，xをC\_yへ分離する作業とです．これはxの所属クラスを表す離散値ラベルyの値を決まるということになります．つまりxから離散値y(x)への写像といえるわけです．

　2つ目は回帰です．先ほどのクラス分類のようにクラスが離散的とは限りません．例えば

，過去数日の気象データｘから明日の気温を予測したいとき，ｙは温度の数値に相当し，これは連続な実数になります．このように，データから，それに対応する実数値を予測する作業のことを回帰と呼びます．これは，与えられたｘを対応するｙに変換するための関数y(x)を決めるということで，xから実数値y(x)への写像といえます．

　回帰におけるタスクの応用例として，機械翻訳，音声認識，異常検知，データ次元削除などがあります．

スライド6　2.2機械学習とは何か

　さきほど，経験Eとはデータ集合＝データセットと説明しましたがこのデータセットについていくつか紹介していきます．

　まず1つ目はMNISTです．アメリカの研究所が提供した手書き数字のデータベースをシャッフルして作られたデータセットです．手書き数字の画像データの集まりで1枚28×28ピクセルの画像になっています．そして，これには訓練用画像に6万枚，テスト用画像に1万枚が用意されています

２つ目はイメージネットです．これは物体カテゴリ認識のためのデータセットで約1400万枚の自然画像からなる巨大なデータベースです．さらにクラスの数は2万枚にも及びます．

　3つ目はサイファー10ですイメージネットと同じく，自然画像からなるデータセットです．MNISTと同じく訓練用画像に6万枚，テスト用画像に1万枚が用意されていて，こぢんまりとした実験を行いたいときなどに用いられます．

スライド7　2.3統計入門

　機械学習ではデータをもとにしてプログラムがいろいろなタスクをこなせるようにさせることが目標ですが，どのようにしたらプログラムはデータからタスクをこなすための知識を学び取れるかを考えなければなりません．データを科学的に分析する際の数理的手法といえば統計学です．実際，機械学習の手法も統計を基礎として構築されているので，統計入門ということで説明していきます．

　まず，用語について確認していきましょう．データ＝データ集合やサンプル，標本とういのはデータ点の集まりからなります．(手書き文字画像認識の例を話す)ここで，注意してほしいことがあり，これらの用語はかなり乱用されています．データ点を略してデータと読んだり，サンプルをサンプルの要素であるデータ点の意味で用いられたりしています．これらの意味については文脈からきちんと判断するようにしましょう．

スライド8　2.3統計入門

　統計ではデータ点の集まりからデータを分析するわけですが，ここでは推定というものについて考えていきます．なにか母集団について知識を得たいとき，母集団の全体のデータを得ることはたいてい不可能です．例えば，飲酒と健康の関係について調べたいとき，母集団は地球上の人類すべてになりますが人類すべてのデータを得ることは不可能ということです．実際にこれについて調べるときには，ランダムに選び取った少人数に対するデータを分析して結果を調べようとします．このように少数のデータから母集団についての知識を獲得することが推定の目標になります．

　推定にはいくつか仮定をする必要があります．まず，統計解析に用いるデータは母集団から無作為に抽出されたものとみなします．それから，母集団の性質はデータ生成確率により，特徴づけられていると仮定します．これはつまり，不確実性のある現象を確率的にモデル化するということです．そして，最後に母集団から抽出されたサンプルというのはデータ生成確率から抽出されているものであると仮定します．

　３つ目についてですが，ある変数xの実現値xが分布Pから抽出されているということを数式でこのように表します．

　このような仮定によって我々は現象を確率的に予測できるようになるというわけです．

スライド9　2.3統計入門

　以上のことから推定とは母集団について知ると話しましたが，言い換えるとデータ生成分布について知るというのがやりたいことになります．もう少しいうと，データ生成分布を特徴づける量のことをパラメータと呼びますが，このパラメータの値を決めるというのが推定においてやることです．

　データ生成分布がガウス分布であった場合を例に考えてみます．ガウス分布というのはこのような数式で表しますが．ガウス分布の形を決めるものは平均値と分散です．つまり，ガウス分布におけるパラメータは平均値と分散の2つということになります．

　実際のデータ生成分布というのはかなり複雑で無数のパラメータを持っています．そのパラメータについて知るというのは現実的ではないです．そのため，通常は実際のデータ生成分布をよく近似できると期待できるパラメータの少ないモデル分布を仮定し，そのモデルパラメータの最適値をデータから推定することを行います．

　このように，パラメータの最適値をデータから推定するアプローチをパラメトリックなアプローチと呼びます．