



東京大学
素粒子物理国際研究センター
International Center for Elementary Particle Physics
The University of Tokyo

コライダー実験における 深層学習

2024年 7月 7日

LCWS プリスクール「将来電子・陽電子コライダーの物理と技術」

東京大学 素粒子物理国際研究センター

齊藤真彦

アウトライン

- 深層学習のイントロダクション
 - 機械学習・深層学習とは？
 - 機械学習・深層学習の基礎知識
 - 分類、学習、パーセプトロン、モデルアーキテクチャ
- コライダー実験への深層学習の応用例
 - 物理データ解析
 - 物理オブジェクト再構成
 - 検出器シミュレーション

機械学習とは？

ChatGPT 4o >



機械学習とは？

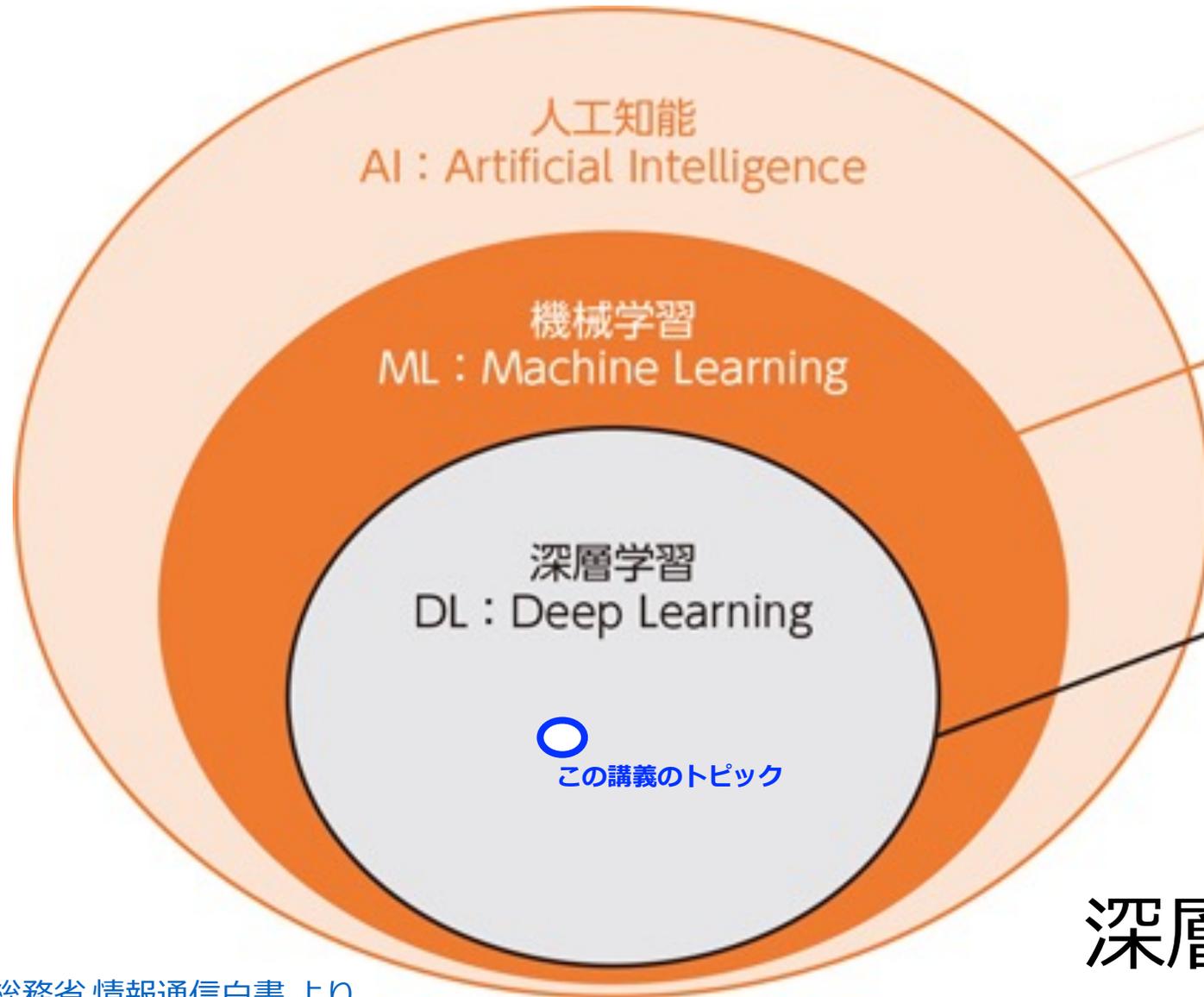
 機械学習（きかいがくしゅう、Machine Learning）とは、コンピュータに大量のデータを与え、そのデータを基にパターンやルールを自動的に学習し、予測や判断を行う技術や方法論のことを指します。以下に、機械学習の基本的な概念と主要な種類について説明します。

基本的な概念

1. **データ**: 機械学習の基盤はデータです。データは、機械学習アルゴリズムが学習し、予測を行うための情報源です。
2. **特徴量**: データの中で、学習に用いる属性や指標のこと

- 機械が**データから学習** (経験)
 - **人間の手作業** ではない
- 経験を基に**タスク(予測等)をより良く実行**

機械学習、深層学習、AI



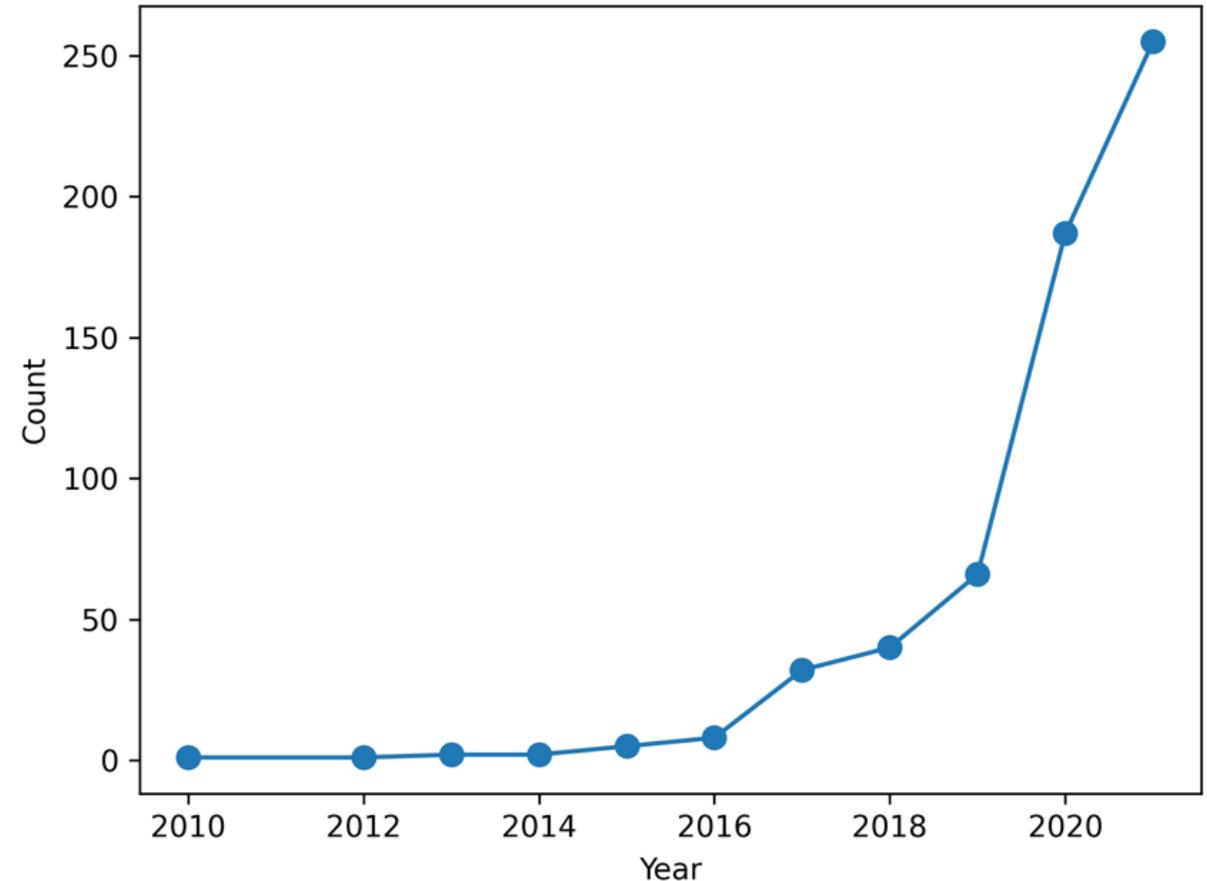
- 人間の思考プロセスと同じような形で動作するプログラム全般
- あるいは、人間が知的と感じる情報処理・技術全般
- AIのうち、人間の「学習」に相当する仕組みをコンピューター等で実現するもの
- 入力されたデータからパターン/ルールを発見し、新たなデータに当てはめることで、その新たなデータに関する識別や予測等が可能
- 機械学習のうち、多数の層から成るニューラルネットワークを用いるもの
- パターン/ルールを発見する上で何に着目するか（「特徴量」）を自ら抽出することが可能

深層学習 ⊂ 機械学習 ⊂ AI

深層学習と高エネ実験

- 1988: [B. Denby, Neural Networks and Cellular Automata in Experimental High-energy Physics](#)
 - “Neural network”が初めて使われた
- 2014: [P. Baldi, et. al., Searching for exotic particles in high-energy physics with deep learning](#)
 - “Deep learning”が初めて使われた
- 2014: [Kaggle HiggsML Challenge](#)
 - 機械学習コミュニティを巻き込んだコンペ
 - 深層学習、BDT等様々な手法が用いられた

[HEPML-LivingReview](#) 登録論文数



ここ10年で深層学習は**当たり前**の技術に

機械学習・深層学習の 基礎知識

機械学習の分類

1. 教師あり学習
2. 教師なし学習
3. 強化学習

機械学習の分類

1. 教師あり学習

HEPでは多くがこれ

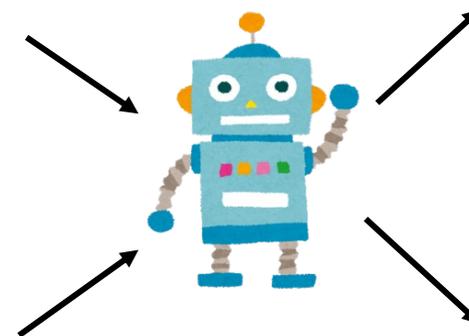
入力(x) と 出力(y) の関係性を学習する

例

$$y = f(x)$$

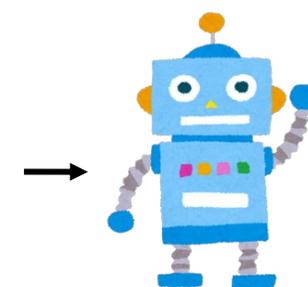
- 写真 (x) に
ネコが写っているか(y) を判定する
- 複数の粒子の運動量 (x) から
Bメソンが生成されたか(y) を判定する
- 検出器の測定値 (x) から
粒子のエネルギー(y) を推定する

入力 (x) \rightarrow モデル ($f(x)$) \rightarrow 出力 (y)



猫? \rightarrow Yes

猫? \rightarrow No



b-jet? \rightarrow Yes

機械学習の分類

HEPでも最近利用例が増加

2. 教師なし学習

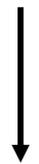
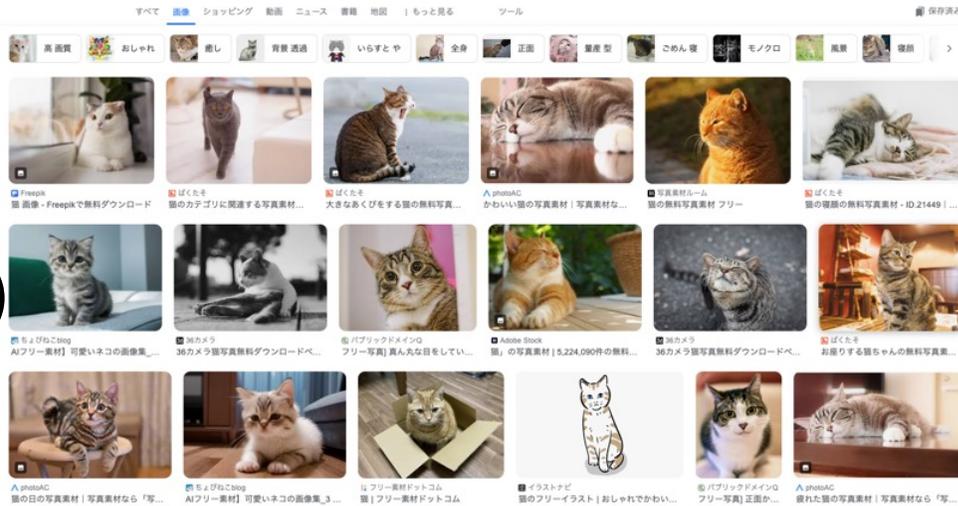
入力(x)のみからその分布や特徴を学習する $p(x)$

- 条件付き(c)の分布 ($p(x|c)$) を学習する場合も

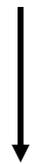
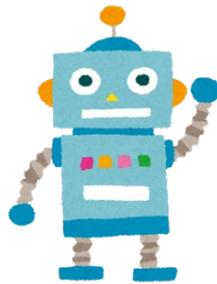
例

- テキスト (c) を与えたら絵 (x) が生成される
- 偽の検出器反応 (x) を生成する。

入力 (x)



モデル ($p(x)$)



サンプリング (x')



(DALL-Eで生成した画像なので、
厳密には教師あり)

機械学習の分類

1. 教師あり学習

HEPでは多くがこれ

入力(x) と 出力(y) の関係性を学習する

例

$$y = f(x)$$

- 写真 (x) に
ネコが写っているか(y) を判定する
- 複数の粒子の運動量 (x) から
Bメソンが生成されたか(y) を判定する
- 検出器の測定値 (x) から
粒子のエネルギー(y) を推定する

- 他にも 半教師あり学習、弱教師付き学習、自己教師あり学習、...
- 境界線が曖昧になっています

2. 教師なし学習

HEPでも最近利用例が増加

入力(x) のみからその分布や特徴を学習する $p(x)$

- 条件付き(c)の分布 ($p(x|c)$) を学習する場合も

例

- テキスト (c) を与えたら絵 (x) が生成される
- 偽の検出器反応 (x) を生成する。

3. 強化学習

たまに見る程度

行動と報酬から最適な方策(Policy)を探索する

例

- 多くの将棋の対局を繰り返すことで最適な指し手を選択できるようになる。

機械学習における学習と目的

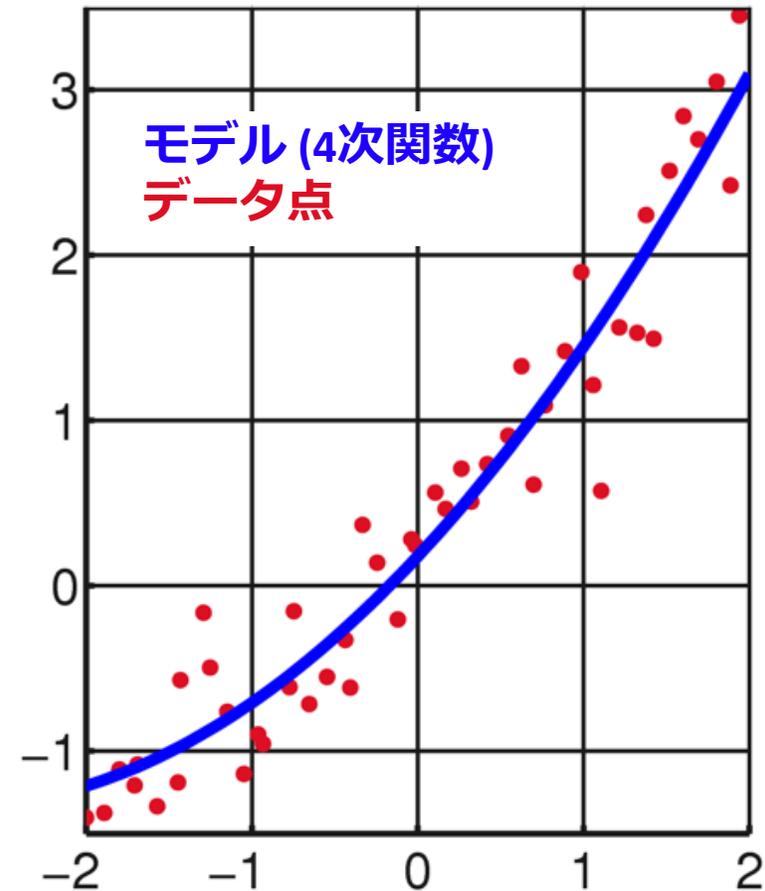
皆さんおなじみの関数フィットとだいたい同じ

最小二乗法

$$J = \sum_{i=1}^n (y_i - f(x_i|\theta))^2$$

を最小化するモデルパラメータ (θ) を探す

[Wikipedia](#)



機械学習における学習と目的

皆さんおなじみの関数フィットとだいたい同じ

ただしモデルが複雑でパラメータ数が非常に多い

最小二乗法

$$J = \sum_{i=1}^n (y_i - f(x_i|\theta))^2$$

を最小化するモデルパラメータ (θ) を探す

機械学習での学習 (最尤法)

$$\mathcal{L} = \sum_{i=1}^n L'(f(x_i|\theta), y_i)$$

を最小化するモデルパラメータ (θ) を探す

最小化する目的関数のことを損失関数(ロス関数)といいます

機械学習における学習と目的

皆さんおなじみの関数フィットとだいたい同じ

ただしモデルが複雑でパラメータ数が非常に多い

最小二乗法

$$J = \sum_{i=1}^n (y_i - f(x_i|\theta))^2$$

を最小化するモデルパラメータ (θ) を探す

機械学習での学習 (最尤法)

$$\mathcal{L} = \sum_{i=1}^n L'(f(x_i|\theta), y_i)$$

を最小化するモデルパラメータ (θ) を探す

注: 真に最小化したい値は**汎化誤差** (=フィットに使うデータ以外も含めた**真の分布に対する誤差**)

(真の分布を使うことはできないので、経験誤差(観測データのみから得られた誤差)で代替している)

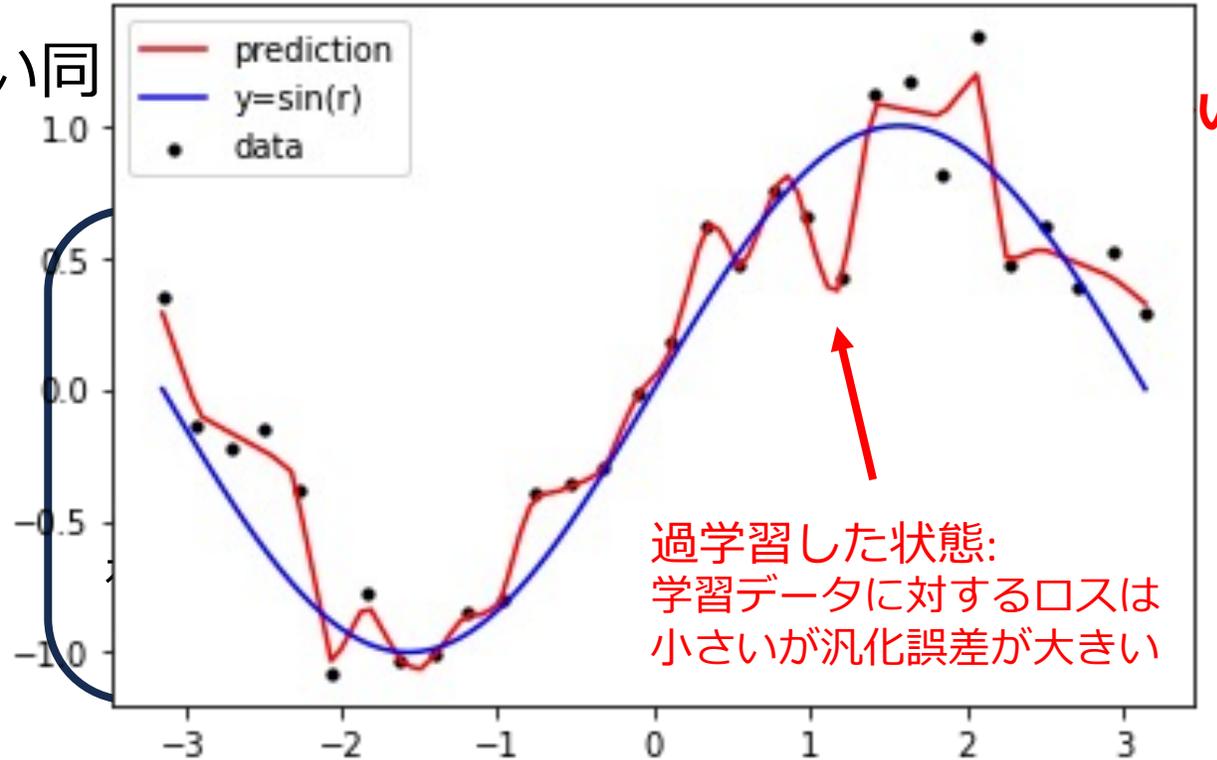
機械学習における学習と目的

皆さんおなじみの関数フィットとだいたい同

最小二乗法

$$J = \sum_{i=1}^n (y_i - f(x_i|\theta))^2$$

を最小化するモデルパラメータ (θ) を探す



注: 真に最小化したい値は**汎化誤差** (=フィットに使うデータ以外も含めた**真の分布に対する誤差**)

(真の分布を使うことはできないので、経験誤差(観測データのみから得られた誤差)で代替している)

深層学習モデルはパラメータ数が非常に大きく過学習を起こしやすいため、
大量の学習データが必要です。

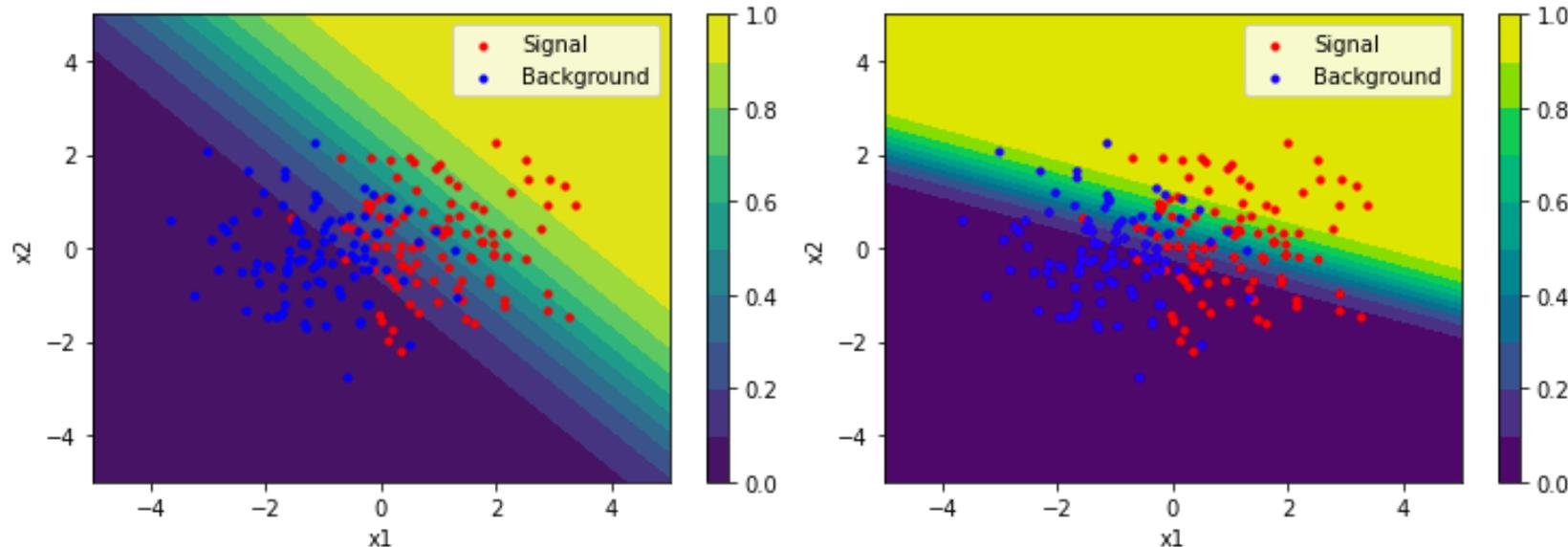
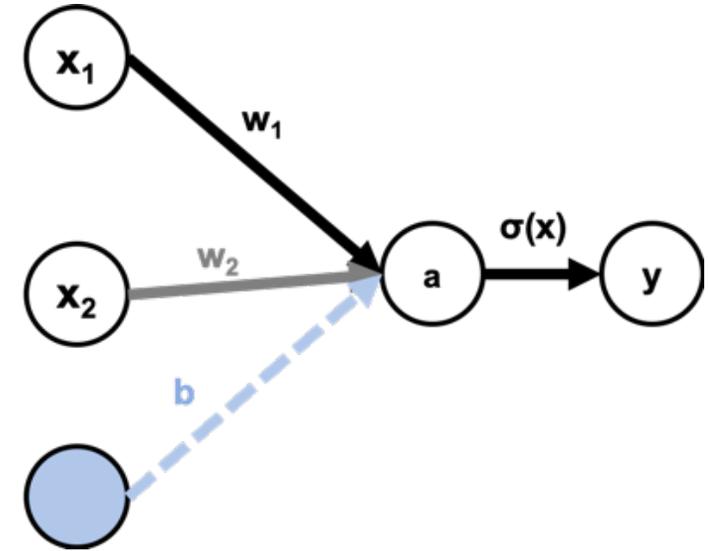
深層学習モデルの基礎(1): パーセプトロン (Perceptron)

1957年に開発されたアルゴリズム

入力(ベクトル)を線形変換

$$y(x|\mathbf{w}, b) = \sigma(\mathbf{w} \cdot \mathbf{x} + b)$$

ベクトルのそれぞれの成分を非線形変換

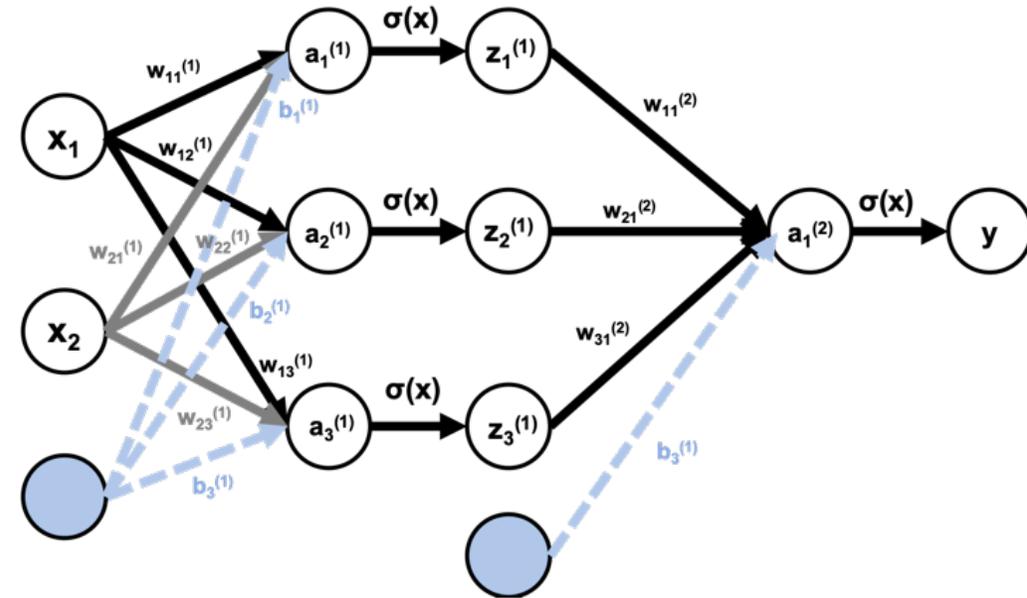


パーセプトロンは領域を直線で分けるような分類をする

線形分離可能な問題を
解くことができる

深層学習モデルの基礎(2): 多層パーセプトロン (Multi-layer Perceptron)

- パーセプトロンを多層にしたもの
- 線形分離できない問題も解ける
- 中間層のノード数が無限に大きければ、
どんな関数も表現できる (万能近似定理)



1層目のパーセプトロン
$$z_j^{(1)} = \sigma \left(\mathbf{w}_{ij}^{(1)} x_i + b_j^{(1)} \right)$$

2層目のパーセプトロン
$$y^{(2)} = \sigma \left(\mathbf{w}_j^{(2)} z_j^{(1)} + b^{(2)} \right)$$

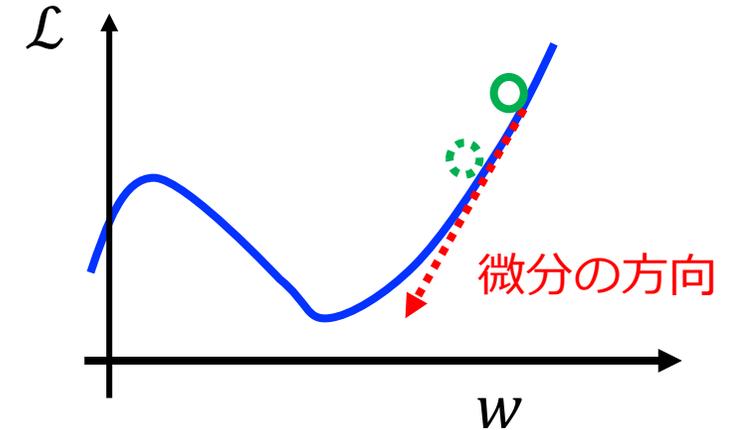
次元数と層数を大きくすることでいくらでも柔軟な変換が(原理上は)可能になる

→ どうやって学習するか?

深層学習モデルの基礎(3): モデルの学習(最適なパラメータの決定)

- 深層学習モデルのパラメータは勾配法で最適化する
 - 目的関数に対する勾配 ($\frac{\partial \mathcal{L}}{\partial w}$) を使って w を少しずつ更新する

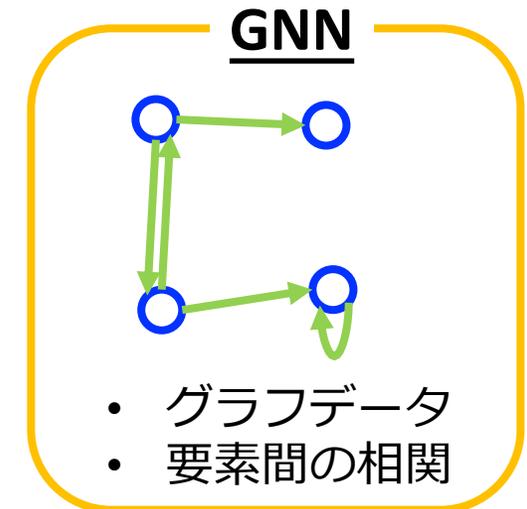
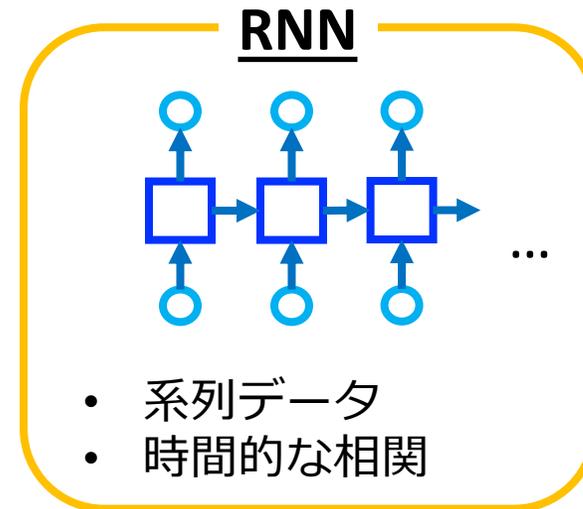
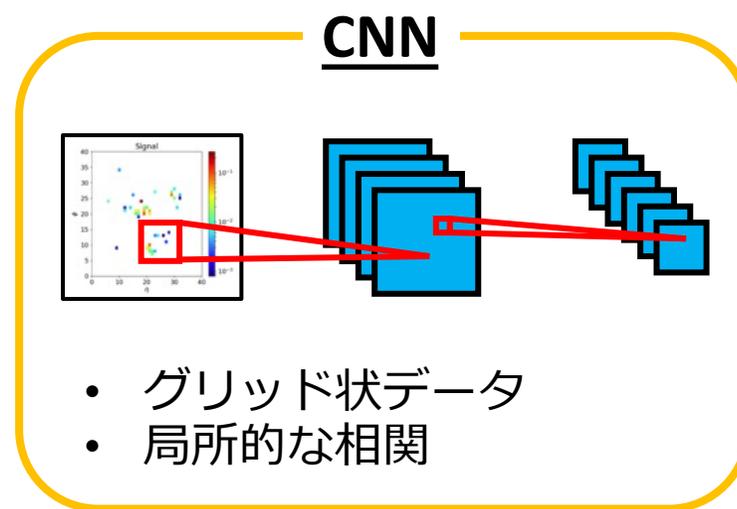
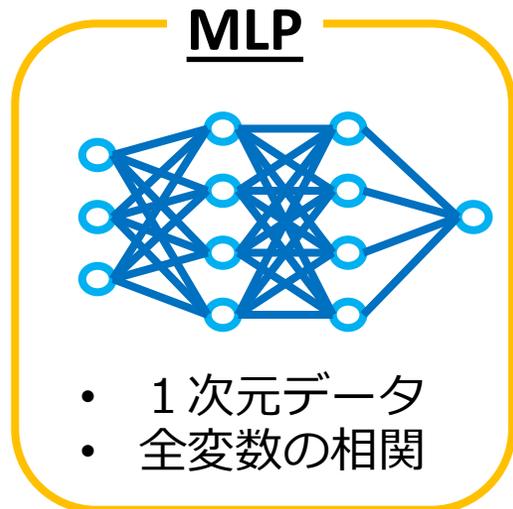
$$w^{(k+1)} = w^{(k)} - \epsilon \cdot \left. \frac{\partial \mathcal{L}}{\partial w} \right|_{w=w^{(k)}}$$



- 効率よく勾配の値を計算する手法 → **誤差逆伝搬法** (backpropagation)
 - パラメータ数 N に対して $O(N)$ で N 個の勾配が計算できる (かつ計算機で計算できる)
 - 深層学習モデルはパラメータがとても多い (e.g. GPT-3 $\sim 10^{11}$) のでこの手法が必須

深層学習モデルの基礎(4): 深層学習モデルアーキテクチャ

- 多層パーセプトロンは(MLP)は何でも近似できる → **過学習**を起こしやすい
学習データだけを記憶してしまう。
汎化誤差は悪い
- 問題の構造に合わせたネットワーク構造
 - 畳み込みニューラルネットワーク (CNN) : グリッドデータ (画像)
 - リカレントニューラルネットワーク (RNN) : 系列データ (自然言語, 音声)
 - グラフニューラルネットワーク (GNN) : グラフデータ (ネットワーク)



コライダー実験への深層学習の応用例

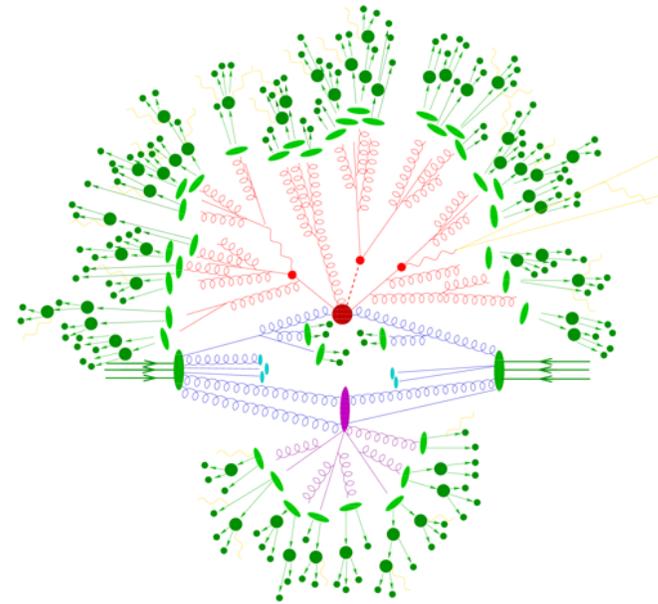
ATLAS実験・ILC実験を中心に

コライダー実験と機械学習

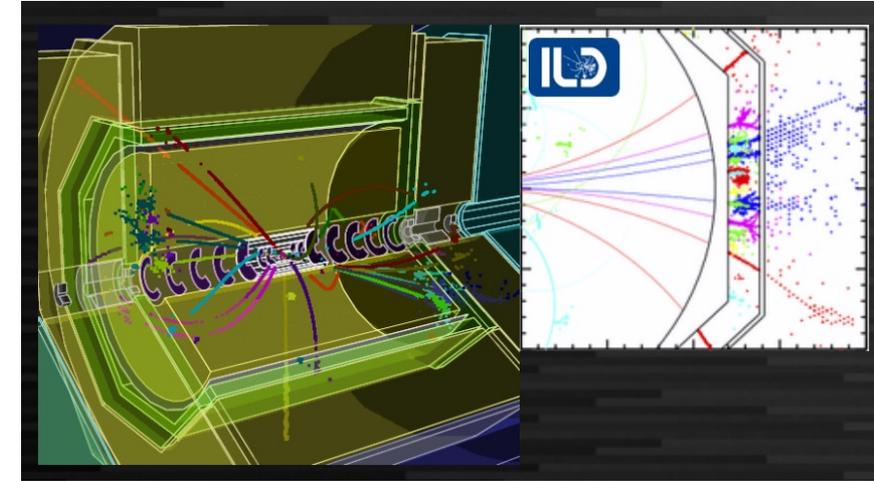
1. 加速器で粒子を加速・衝突させる



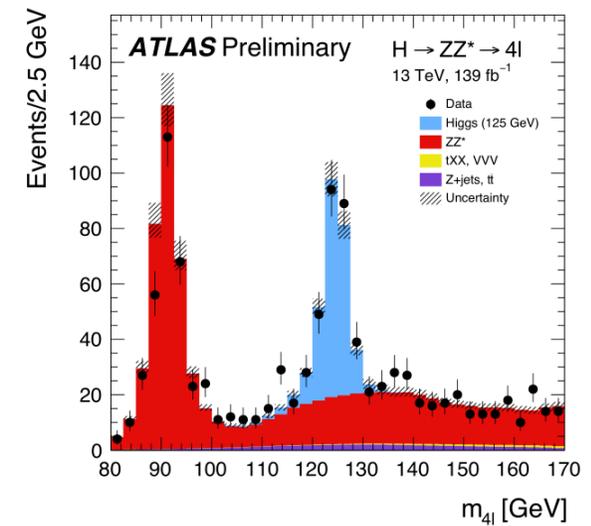
2. 多数の粒子が生成される



3. 生成された粒子を検出し、再構成する



4. 実験データと理論を比較



- 機械学習・深層学習はいたるところに使われ(始め)ている
 - ピックアップしていくつかの応用例を紹介します。

事象選別と深層学習(1): BDT vs DNN

ARTICLE

Received 19 Feb 2014 | Accepted 4 Jun 2014 | Published 2 Jul 2014

DOI: 10.1038/ncomms5308

Searching for exotic particles in high-energy physics with deep learning

P. Baldi¹, P. Sadowski¹ & D. Whiteson²

- 事象選別への深層学習の応用は最もシンプルな応用例
- 以前から機械学習が使われている分野
 - 典型例: **高レベル特徴量**を基に BDT で分類

人間が(物理知識を使って)設計した変数: 例: m_{bb} (Higgs由来なら $\sim m_h$)、 m_{jbb} (top由来なら $\sim m_{top}$)

- 深層学習は**低レベル特徴量**を活かすことで性能向上できる

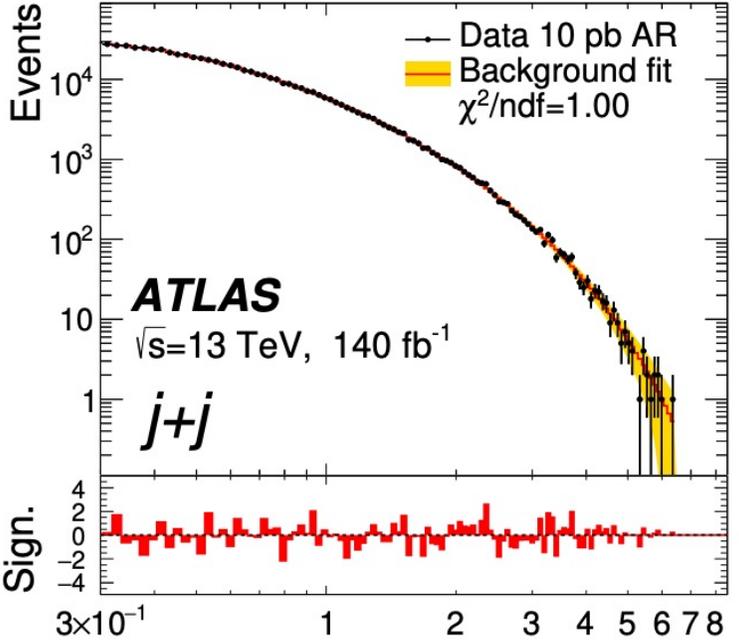
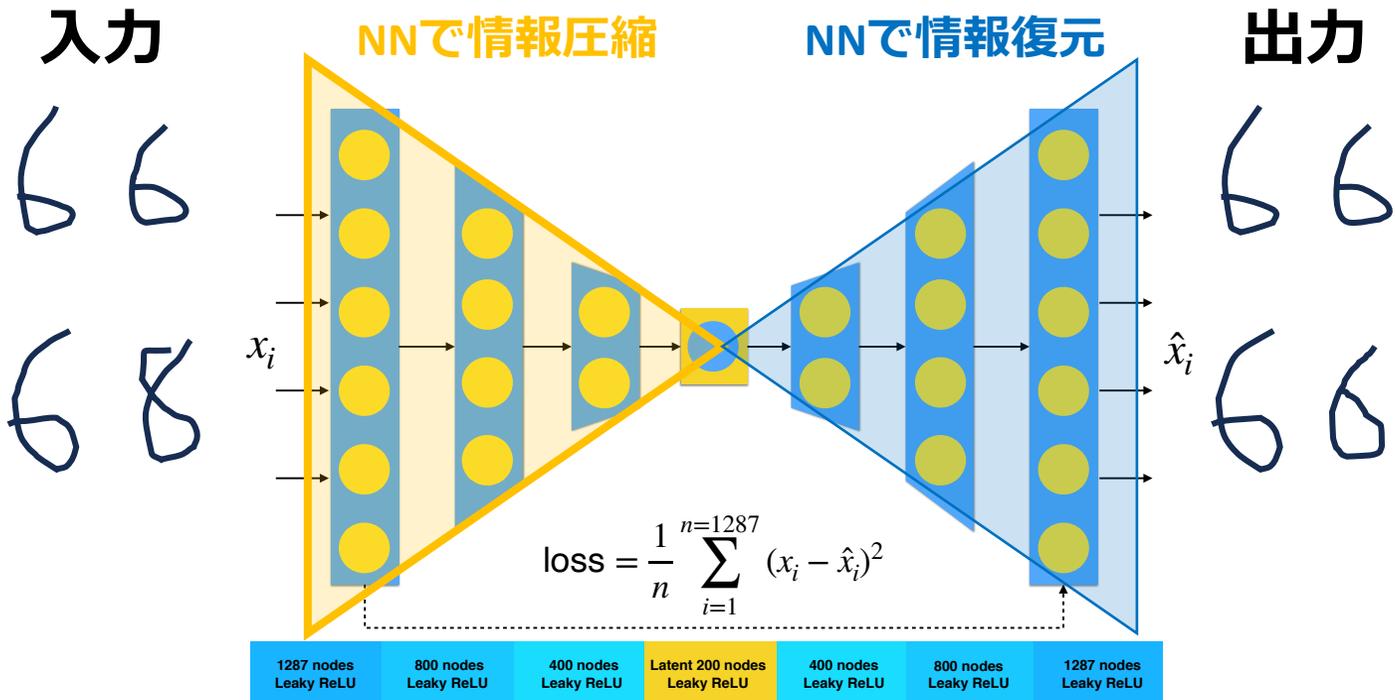
生データに近い特徴量: 例: 測定粒子(ジェット, 電子, ...) 運動量, 粒子ID

Technique	低レベル特徴量	高レベル特徴量	AUC (AUC: 高いほど性能良)
	Low-level	High-level	Complete
BDT	0.73 (0.01)	0.78 (0.01)	0.81 (0.01)
NN	0.733 (0.007)	0.777 (0.001)	0.816 (0.004)
(深層学習) 5層MLP DN	0.880 (0.001)	0.800 (< 0.001)	0.885 (0.002)

(今では、より高級なモデルでより低レベル特徴量を使うこともできます)

事象選別と深層学習(2): 異常検知

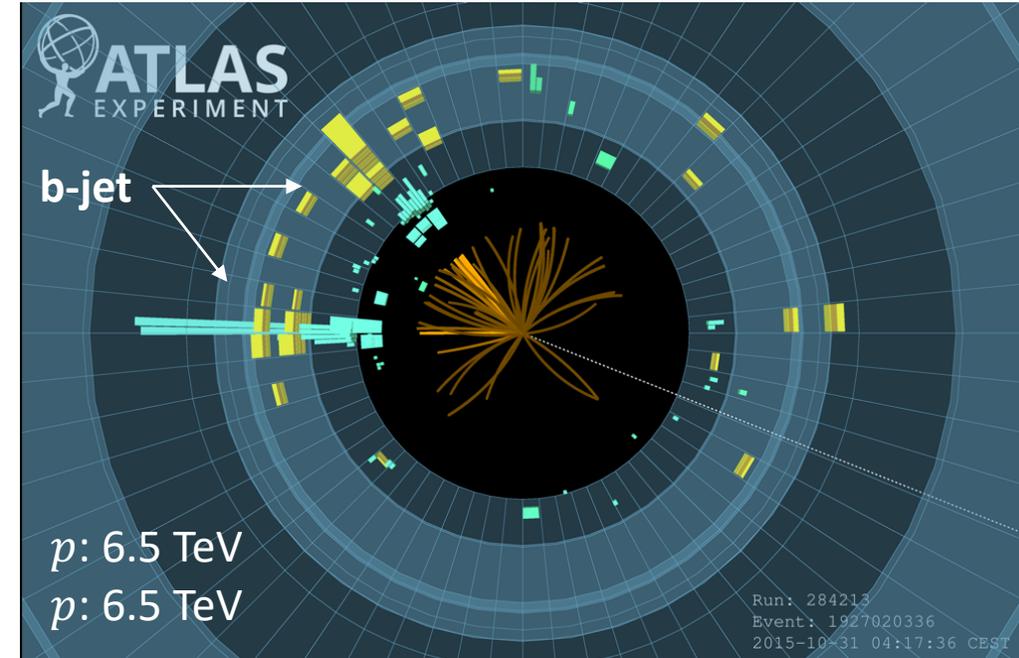
- 最近は**異常検知**の技術を使った事象選別も活発に
- オートエンコーダでデータの中の異常な事象を抽出する
 - 情報の圧縮→復元をすることで、真に重要な情報だけが抽出される
 - 異常データが入力されると、正しく復元されない → これを事象選別に使う



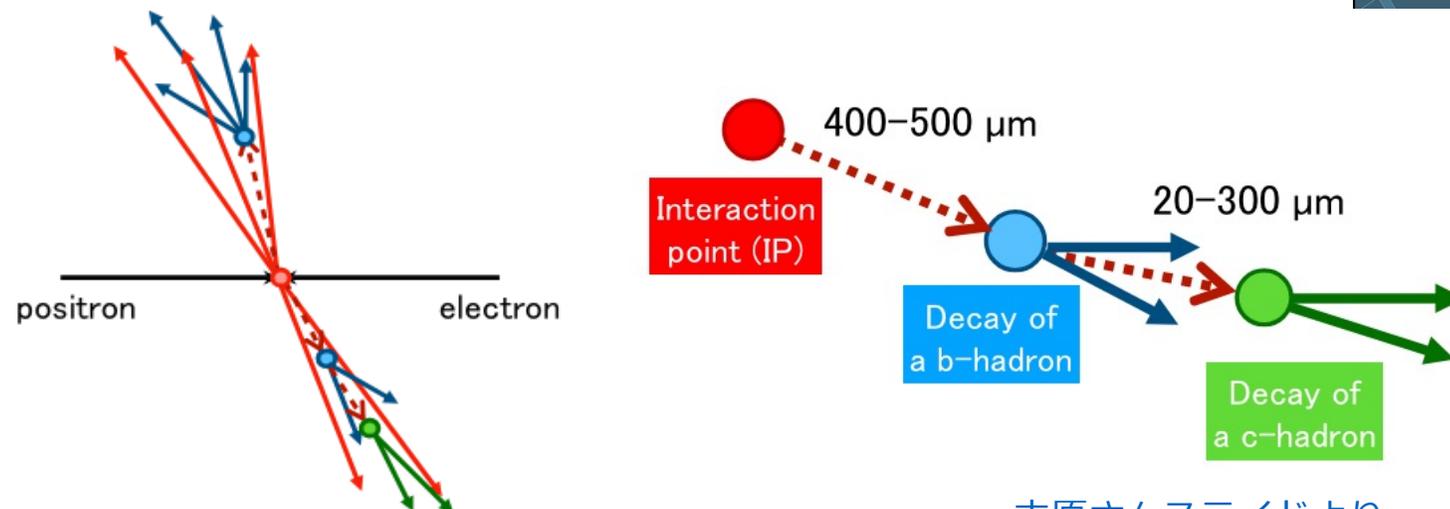
ジェットフレーバー識別と深層学習

- クォークが生成されるとジェットとして観測される
粒子の束
- ジェット中の粒子を見てクォークの種類を識別を行う
 - e.g. b/cジェットはジェット中の粒子が衝突点から少し離れたところから現れる
→ 崩壊点検出が重要な入力の一つとなる(次ページ)

ATLAS: $H(Z) \rightarrow b\bar{b}(vv)$ [ATLAS](#)



Decay vertices in a typical event

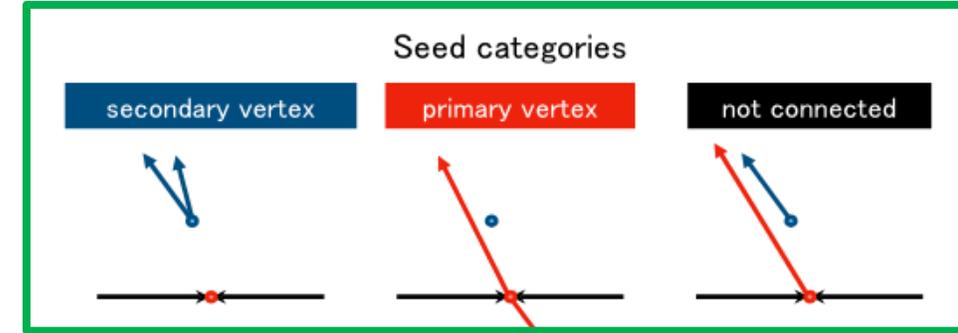


[末原さんスライドより](#)

ジェットフレーバー識別と深層学習: ILC Vertex finding

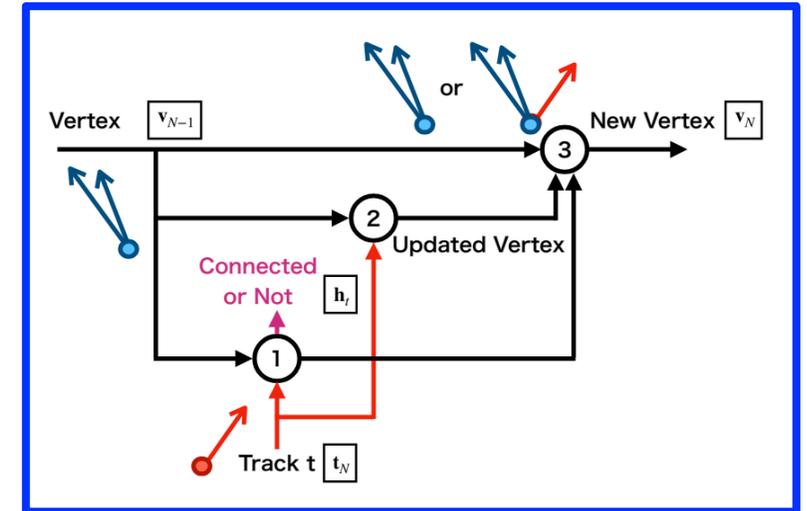
K. Goto et al. (2023)

- 深層学習で崩壊点検出
- 2つの深層学習モデルを組み合わせる
 1. 飛跡のペアから(二次)崩壊点候補を作成 (MLP)
 2. 崩壊点候補に関連する他の飛跡を加える (RNN)
- 飛跡の数は事象ごとに異なるためMLPは向かない



結果

Algorithm	Track origin	Primary	Bottom	Charm	Others
	Total number of tracks	307 657	187 283	180 143	42 888
深層学習	Tracks in secondary vertices	2.2%	63.3%	68.4%	9.5%
DL-based	...from the same decay chain	-	62.3%	67.2%	-
(this work)	...from the same parent particle	-	38.1%	36.2%	6.4%
	Tracks in secondary vertices	0.2%	57.9%	60.3%	0.5%
LCFIPlus	...from the same decay chain	-	57.5%	59.9%	-
(従来手法)	...from the same parent particle	-	34.0%	37.2%	0.3%

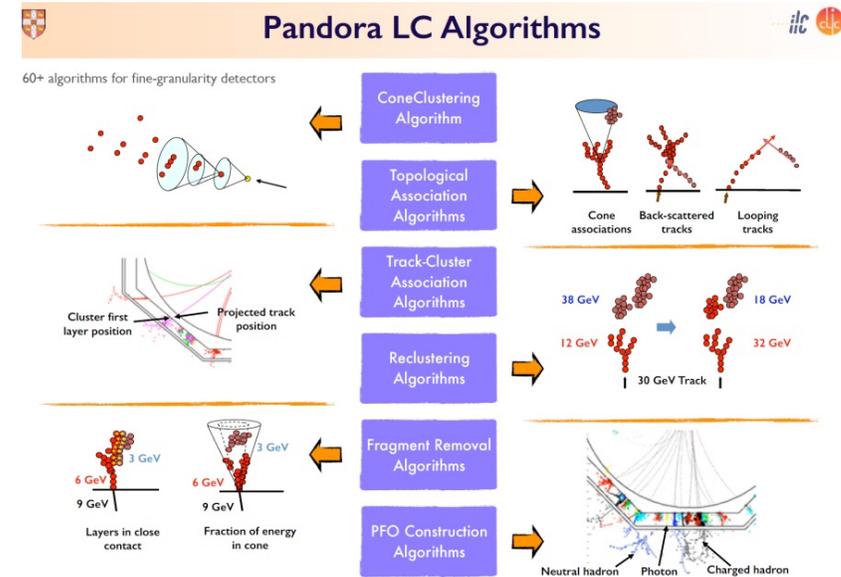
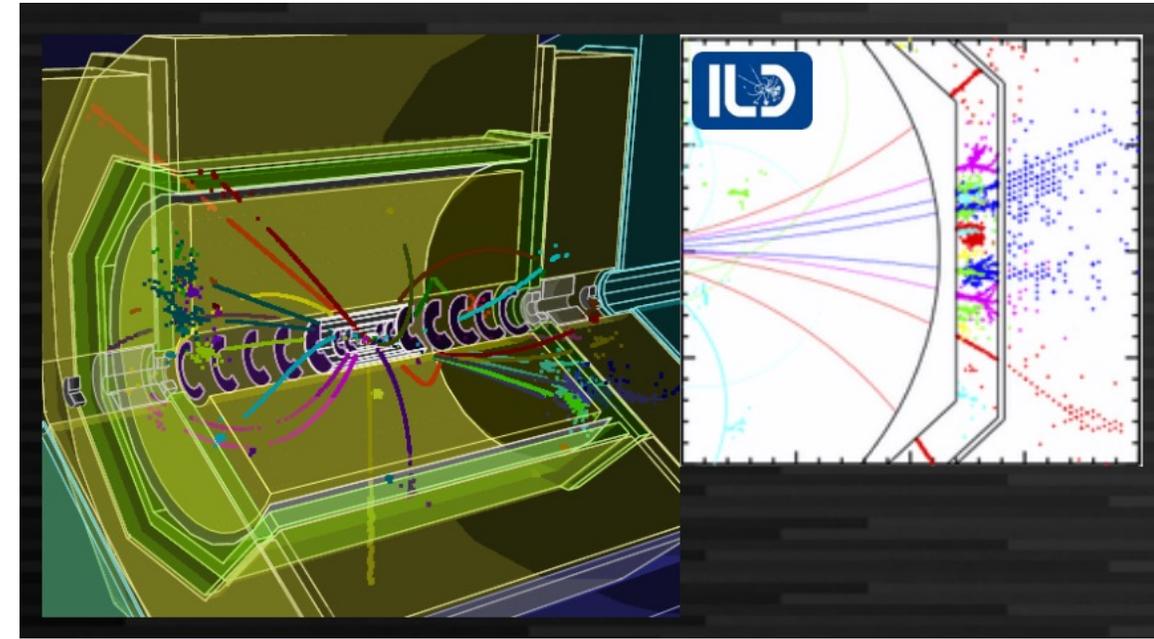


(最近はLLMでも良く知られているTransformerを用いたフレーバー識別も活発に研究されています)

ParticleFlow と 深層学習

Particle Flow アルゴリズム

- 複数の検出器のセンサー情報を組み合わせて、ジェット
トの再構成を行う
- 複数のステップから構成される
 - クラスタリング: 近いセンサー反応点をまとめる
 - マッチング: 異なる検出器間でオブジェクトの統合
 - キャリブレーション: 元の粒子の運動量を再現
 - etc.
- 深層学習で扱うのには**難しいタスク**
 - 多様(tracker/calorimeter)・
多数(e.g. 10^8 sensors for ILD ECAL)の入力
 - 入力と出力の要素数が一致しない
 - いくつかの深層学習モデルが提案されている



ParticleFlow と 深層学習: 様々なモデル

- [GravNet](#) + [Object Condensation](#)

- 同じ粒子由来のヒットが**凝縮**するようにロス関数を定義し、**GNN**で元粒子を再構成

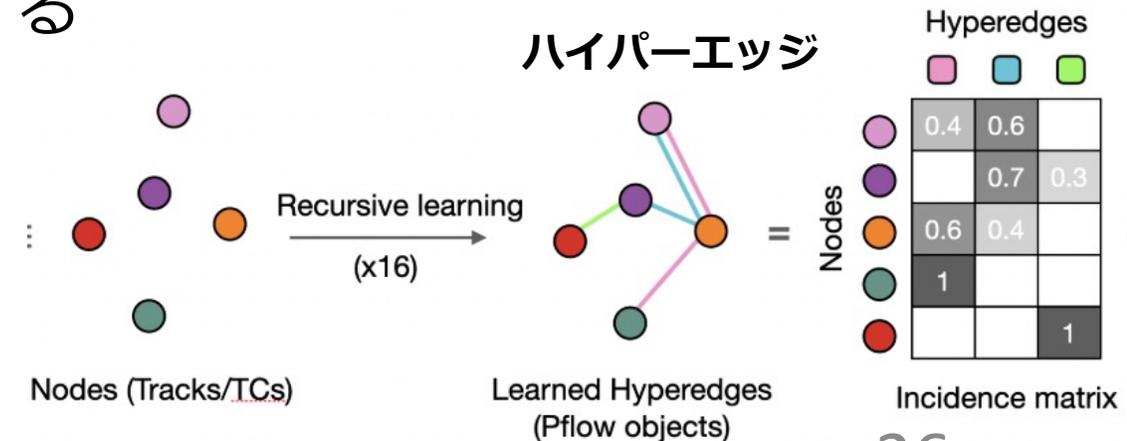
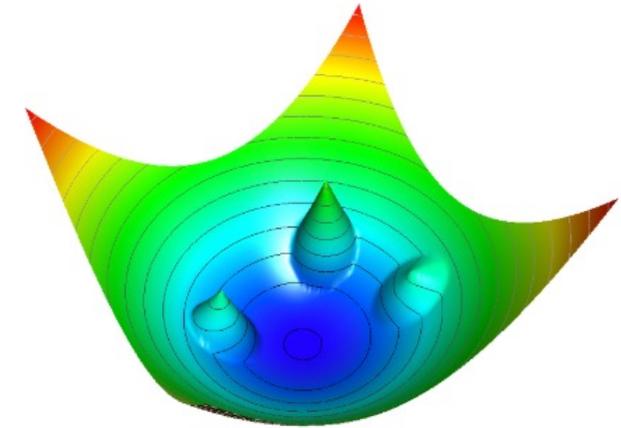
- [ハイパーグラフ](#)

- 複数のノード(ヒット)につながる**ハイパーエッジ**を使って
同じ粒子由来のヒットをクラスタリングする

クラスタリング・マッチングのための深層学習

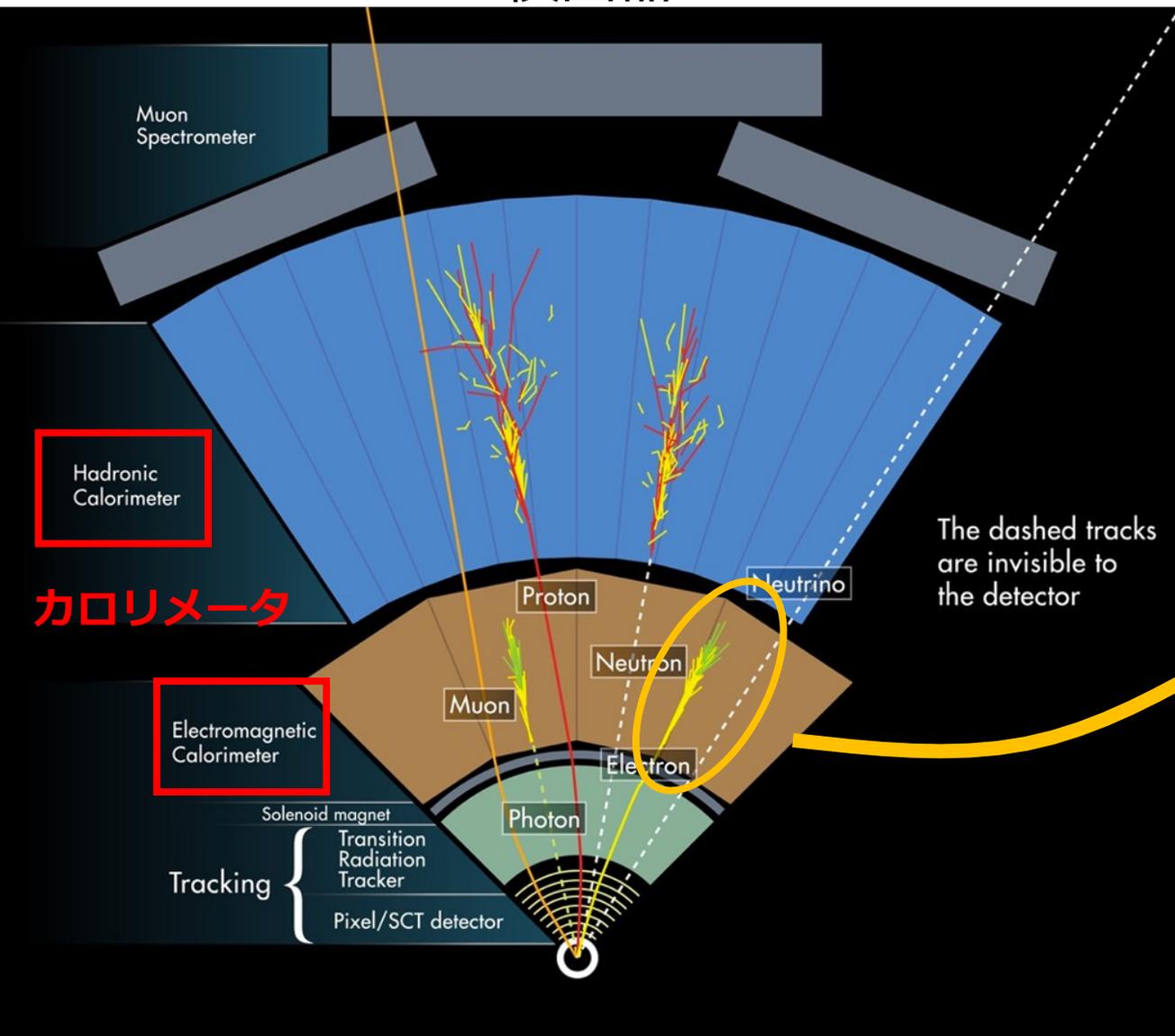
はまだまだ改善の余地あり！

Object condensation損失関数で
用いるポテンシャル関数



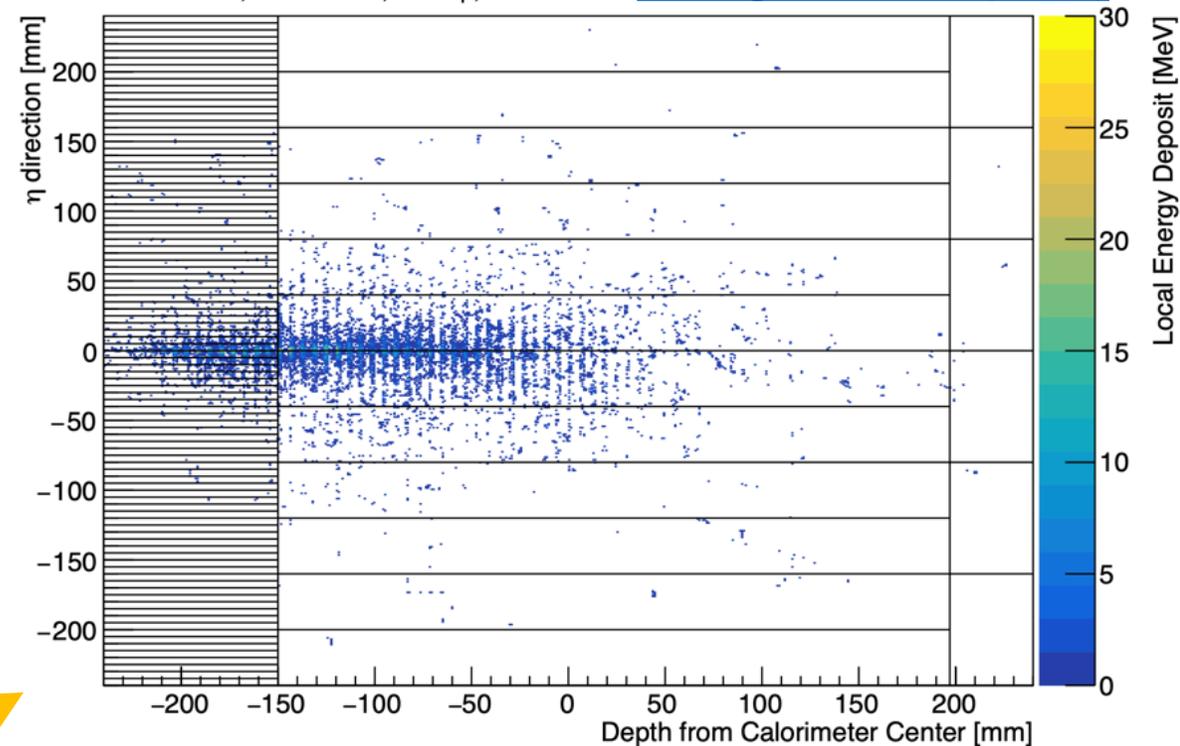
シミュレーションと深層学習

検出器



粒子のシャワー

Geant4, Pb Absorber, IAr Gap, 10 GeV e^- [M. Paganini et al. \(2018\)](#)

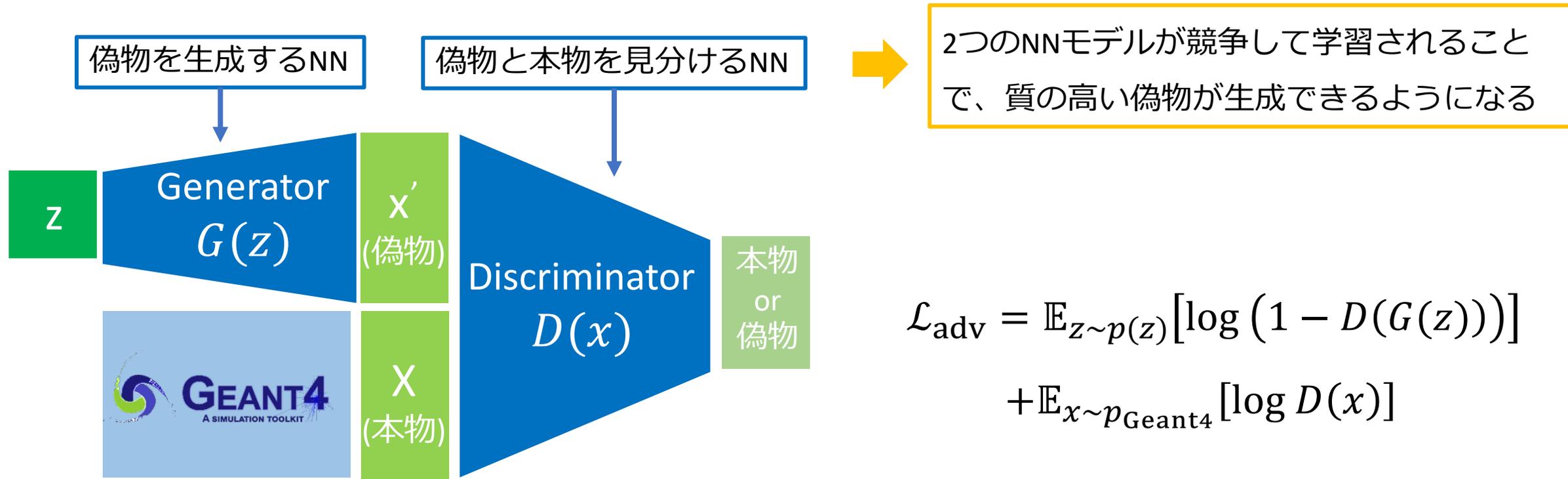


- カロリメータで粒子を止めてエネルギーを測定
 - 入射粒子が多数の粒子に崩壊：シャワー
- カロリメータ中のエネルギー分布をシミュレート
 - Geant4 では一つ一つの粒子を追うため高コスト

シミュレーションと深層学習: CaloGAN(生成モデル)

- 敵対的生成ネットワーク (GAN) で本物 (Geant4) のっぽいシャワーを生成
- 生成時間: Geant4: **6 ~ 126** 秒 (エネルギー依存あり)
GAN : **0.07** 秒

[M. Paganini \(2018\)](#)
[ATLAS Collaboration \(2022\)](#)



$$\mathcal{L}_{\text{adv}} = \mathbb{E}_{z \sim p(z)} [\log (1 - D(G(z)))] \\ + \mathbb{E}_{x \sim p_{\text{Geant4}}} [\log D(x)]$$

(他にもVAE/Normalizing FlowやDiffusion Modelを使った研究も)

おわりに

- 深層学習(+機械学習)はコライダー実験を支える重要な技術になっています。
 - 今後の検出器の複雑化・データ量の増加によってますます重要になるでしょう
- 深層学習の概要と、コライダー実験への応用例を紹介しました
 - それぞれ時間の関係で紹介できなかったことがたくさんあります
- これから深層学習を使う方へ
 - 深層学習(機械学習)が何をやっているかがわかるようにしてください
 - 深層学習の応用はまだまだ広がっていて、できることはたくさんあります