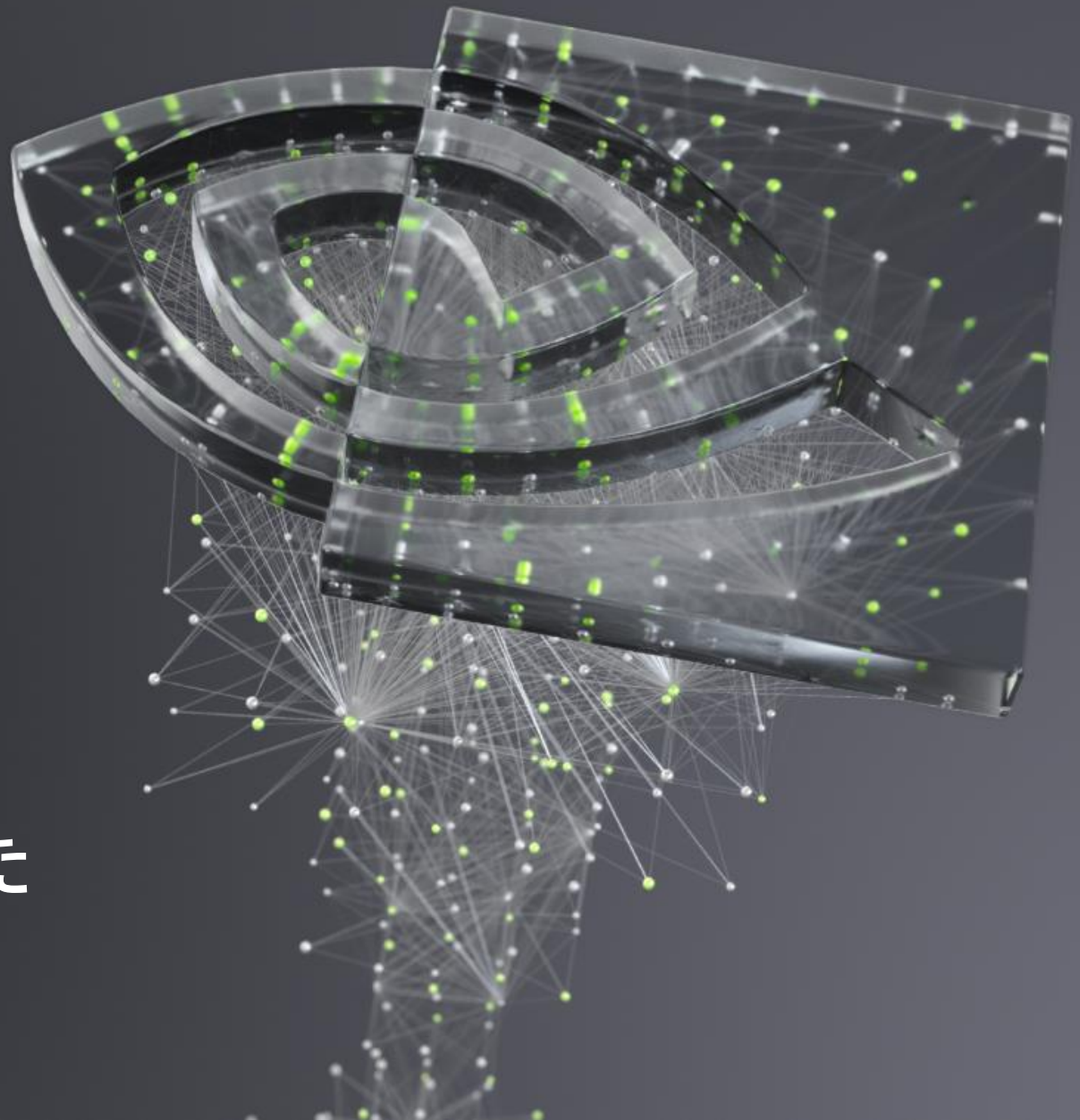
CDLE LT会 #3

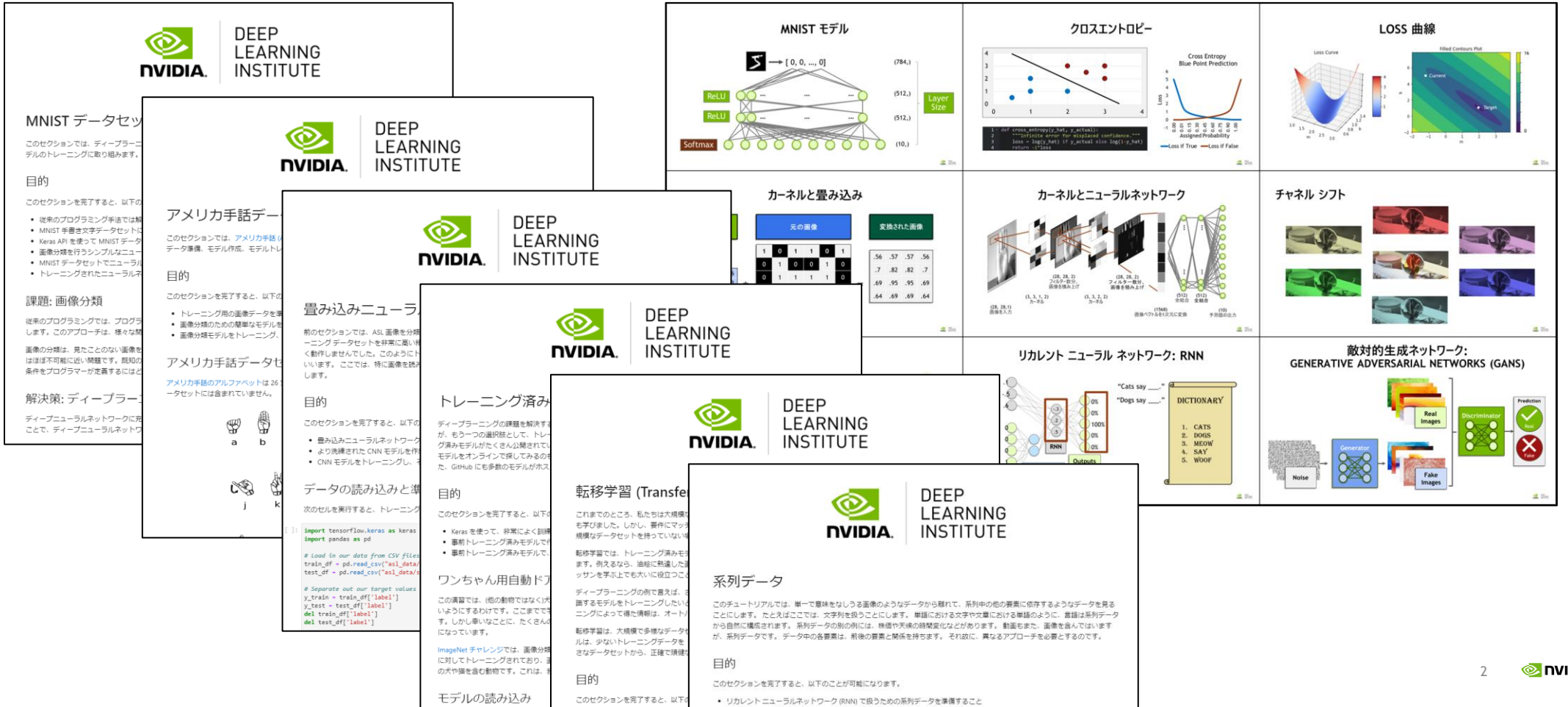


NVIDIA でディープラーニング ハンズオンの講師をやってみた

佐々木邦暢 (@_ksasaki)
エヌビディア合同会社

NVIDIA DEEP LEARNING INSTITUTE (DLI)

今月の GTC で「ディープラーニングの基礎」を日本語で開催しました



ハンズオンのプラットフォーム

- GPU 搭載のクラウドサーバー
- JupyterLab プラットフォーム
- Jupyter ノートブックで対話的なコーディング

NVIDIA DLI 講師と TA 陣

佐々木邦暢
NVIDIA

山崎和博
NVIDIA

村上真奈
NVIDIA

大串正矢
NVIDIA

石関利英
NVIDIA

私に加えて複数の講師と TA が参加者をサポート
この他、優秀な学生 TA が 6 人

DLI 参加者のレベルは様々

私が目指したのは

5年前の私自身に
ディープラーニングを教える

※ 今日は、「初心者のはまりどころ」をテーマにしようかと思ったのですが、それも参加者に失礼なので、恥を忍んで過去の自分を題材にすることにしました。

ディープラーニングの解説を初めて読んだときの話

「単純パーセプトロンは XOR を表現できない」

あー！
線形分離可能な
問題しか解けないのね

(とはならない)

(どこから論理回路出てきた？)

偽らざる私の感想

誤差逆伝播は合成関数の微分の連鎖律で…

あー！なるほどね。

(とはならない)

(連鎖律ってなんだっけ)

偽らざる私の感想

そもそも機械学習ってなに

(私自身がそうであったように)
日本の IT エンジニアは、機械学習を専門的に学んでいないことが珍しくない
そのような方が受講されていることを念頭に、今回のハンズオンを組み立てました

機械学習の例

Iris データセットによるアヤメの分類

がく長	がく幅	花弁長	花弁幅	種別
5.1	3.5	1.4	0.2	I. setosa
4.9	3.0	1.4	0.2	I. setosa
4.7	3.2	1.3	0.2	I. setosa
7.0	3.2	4.7	1.4	I. versicolor
6.4	3.2	4.5	1.5	I. versicolor
6.9	3.1	4.9	1.5	I. versicolor
6.3	3.3	6.0	2.5	I. virginica
5.8	2.7	5.1	1.9	I. virginica
7.1	3.0	5.9	2.1	I. virginica

f (x1, x2, x3, x4)

分類モデル

Iris setosa

Iris versicolor

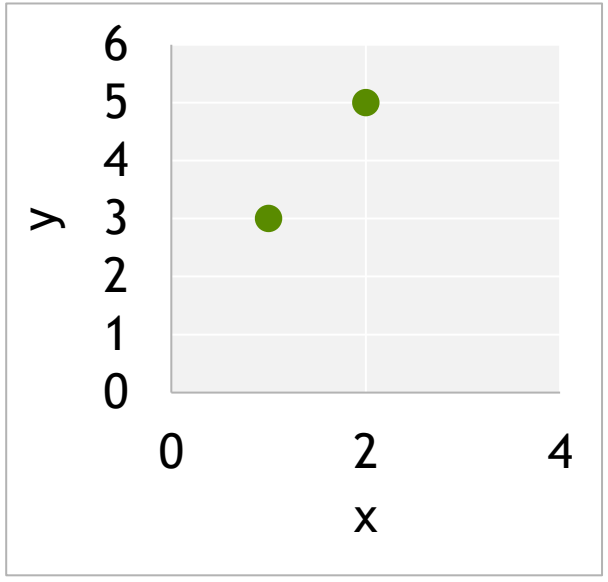
Iris virginica

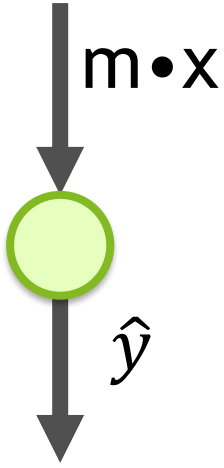
https://en.wikipedia.org/wiki/Iris_flower_data_set

単純なモデル

$$y = mx + b$$

x	y
1	2
3	5





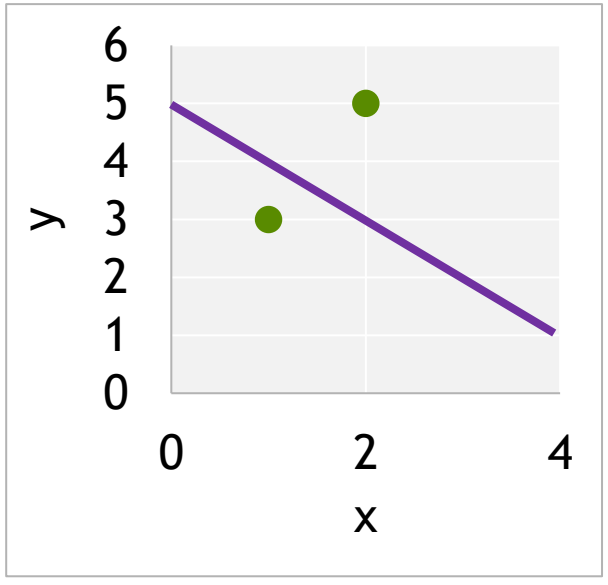
$m = ?$

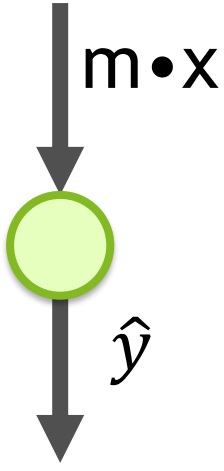
$b = ?$

単純なモデル

$$y = mx + b$$

x	y	$\hat{y}$
1	3	4
2	5	3





Start
Random

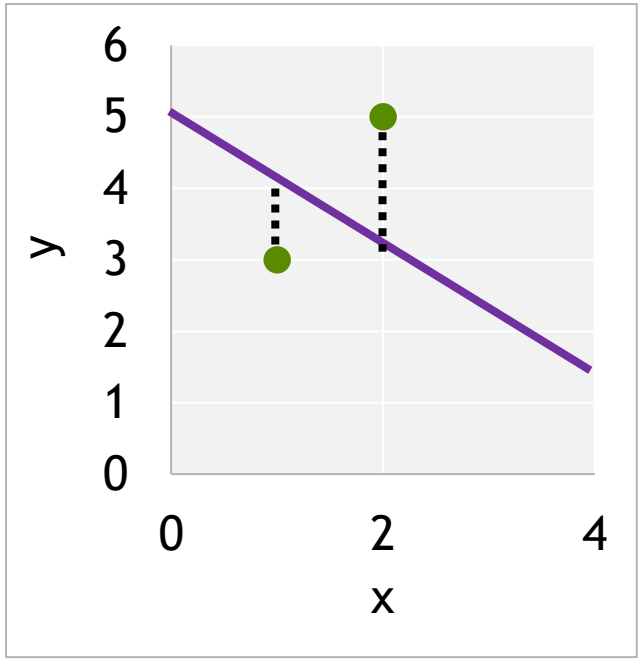
$$m = -1$$

$$b = 5$$

単純なモデル

$$y = mx + b$$

x	y	$\hat{y}$	err^2
1	3	4	1
2	5	3	4
		MSE =	2.5
		RMSE =	1.6



$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y})^2$$

平均二乗誤差

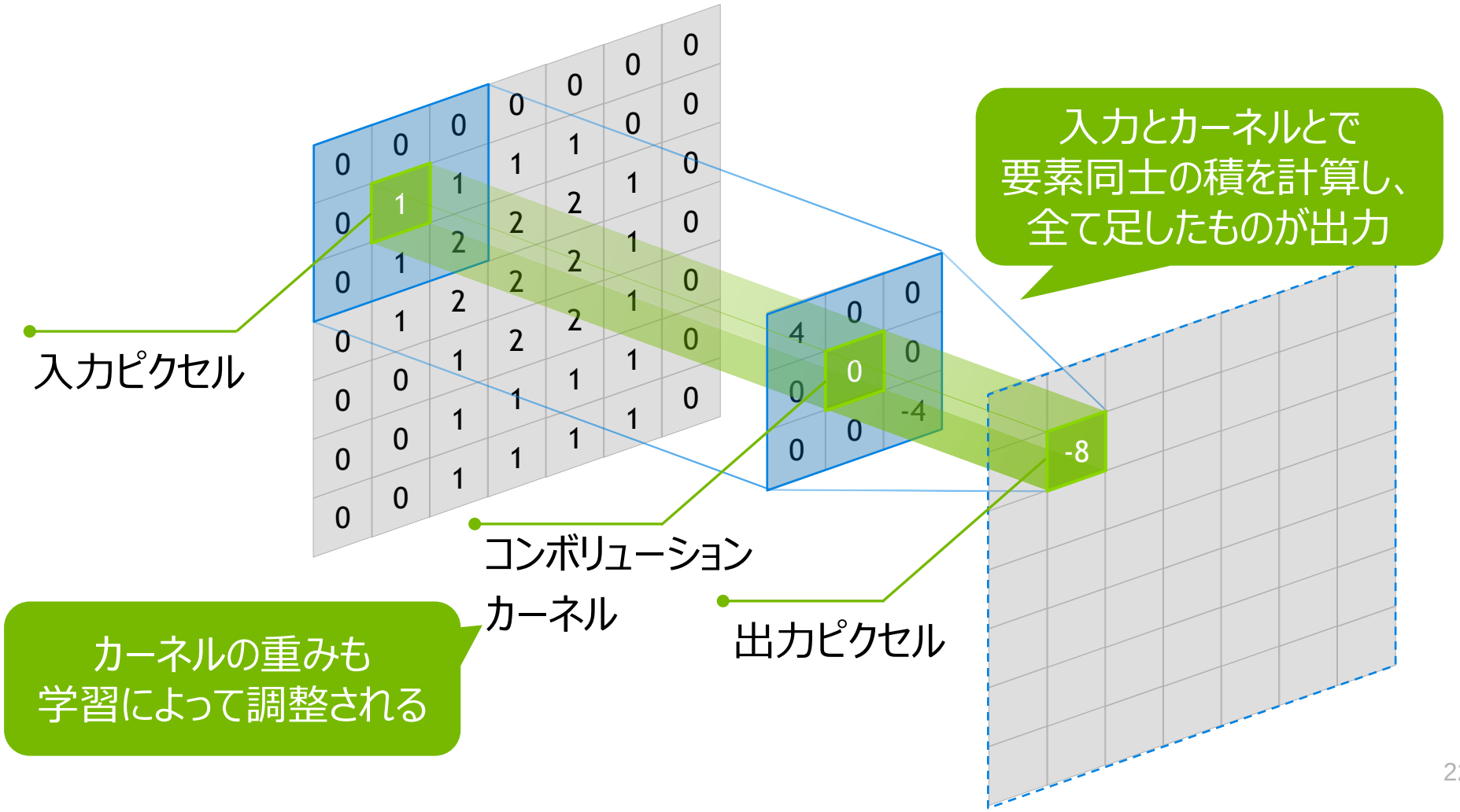
$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(y - \hat{y})^2}$$

二乗平均平方根誤差

わかってしまうとなにが
わからなかったのかわからない

例えば私は、「畳み込みカーネルの個々の要素が学習可能なパラーメーターである」ことがよくわかりませんでした。
どうわからなかったのか、もうよくわかりませんが、わからなかったということだけを、かろうじて覚えています。

CNN: 畳み込み層


入力とカーネルとで
要素同士の積を計算し、
全て足したものが出力
入力ピクセル
コンボリューション
カーネル
出力ピクセル
カーネルの重みも
学習によって調整される

実際にいただいた質問

One-Hot エンコーディング

もともと y_train には 0～9 の数字が入っていると思うのですが、エンコーディング後は0、1になっています。これは何を意味しているのでしょうか？

もともとの 0～9 のカテゴリでも object データではなく数値データなので変換する必要はあるのでしょうか？

実際にいただいた質問

再現性

講師の方の結果と微妙に異なる結果が出ましたが、そんなに再現性がないものなのでしょうか？

実際にいただいた質問

勾配降下法

同じデータを使っているはずなのに epoch が進むにつれて学習が改善されるのはなぜでしょうか。前の epoch の結果を次の学習に使っているということでしょうか。

寄せられた一つ一つの質問が貴重なフィードバックです

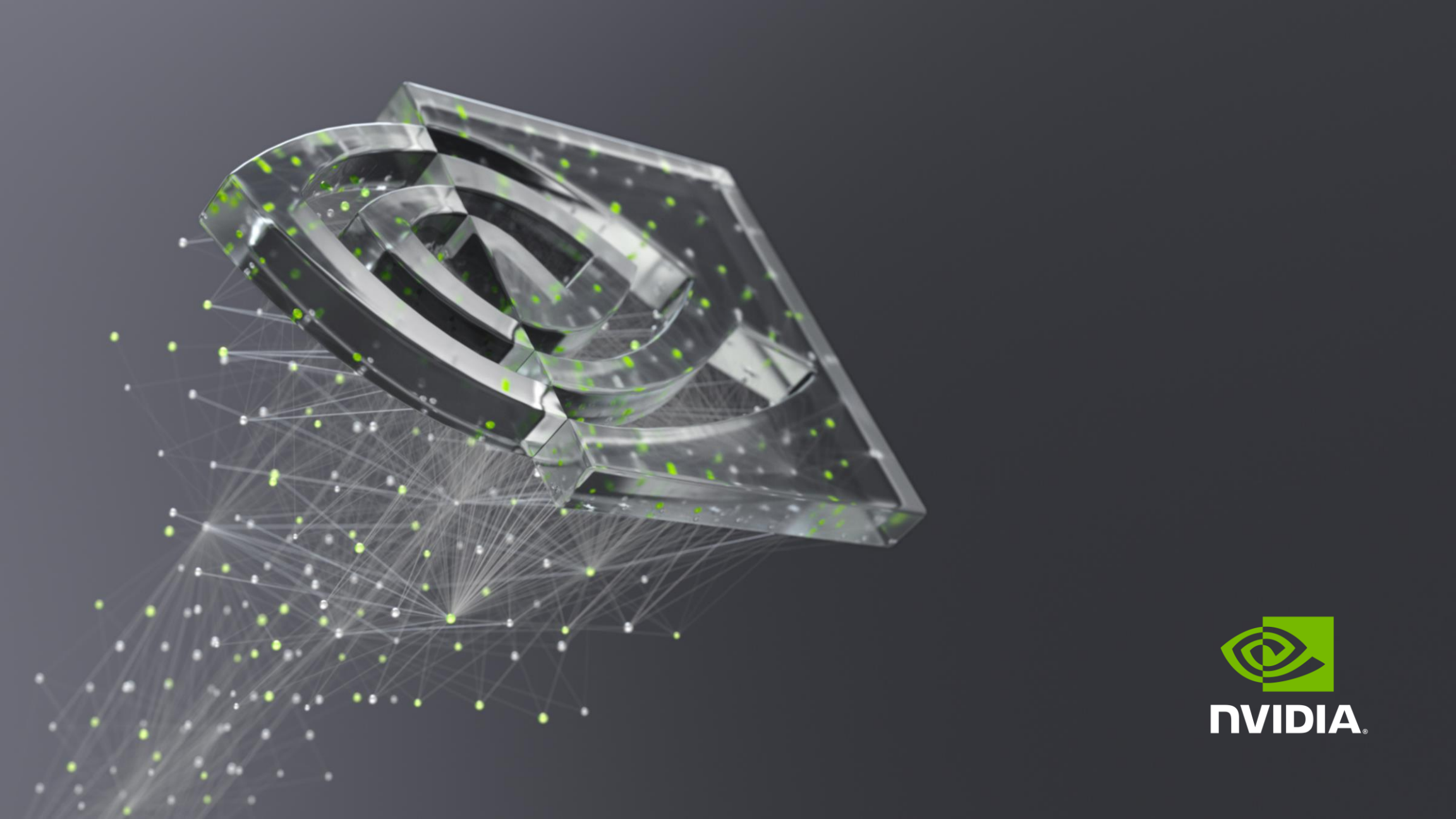
さらにわかりやすく解説するように努めます

ところで、DLI ではいつも、機械学習を専門とする学生さんに TA を頼んでいます。
オンラインのハンズオンでは、彼らが次々に参加者の質問を解決していく様子が
チャットのログなどで可視化されるので、ちょっと感動しました。
(物理イベントでは、会場の各所に散らばった TA が参加者となにを話しているのか、把握できないので)

頂いた感想で一番嬉しかったもの

「楽しみながら取り組むことができました。ありがとうございました！」

また、こういうお言葉をいただけるように頑張ります


NVIDIA