Getting Started with Vision AI Model Training

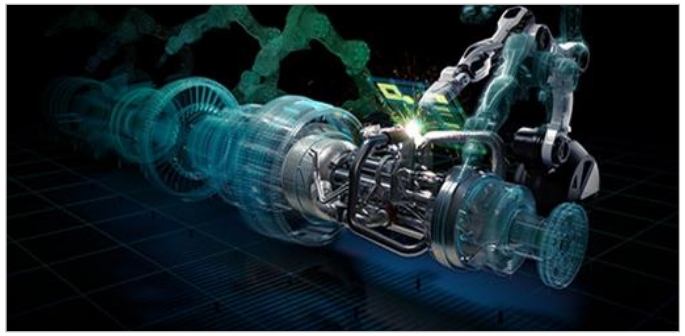
Ekaterina Sirazitdinova
NVIDIA

NVIDIA

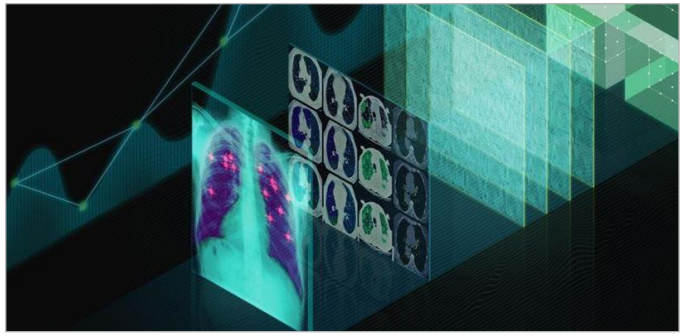
Agenda

- Introduction to deep learning training
  - How it works and how neural networks learn
- What affects training
  - Important training parameters and problem mitigation
- Data preparation
  - Essential steps for creating a deep learning data set
- Where to start
  - Deep learning frameworks and community resources

AI & Deep Learning are Changing the World

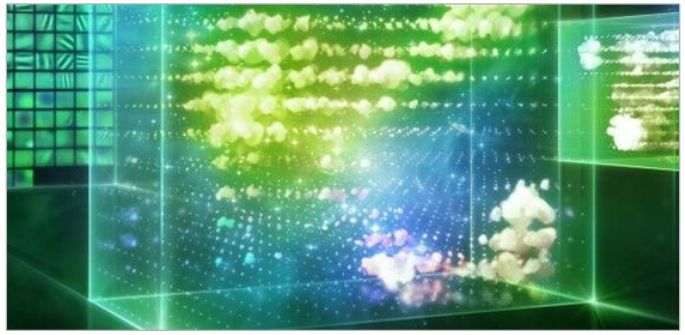


Robotics
Manufacturing, construction, navigation



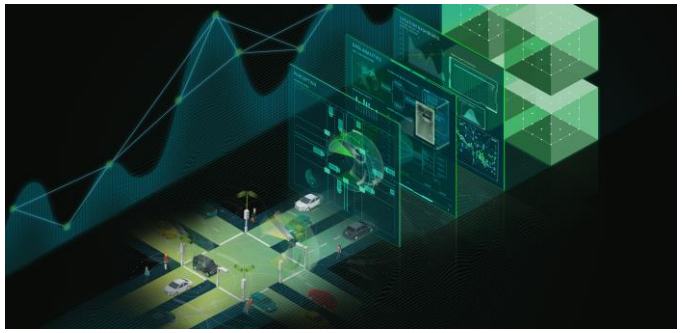
Healthcare
Cancer detection, drug discovery, genomics

Autonomous Vehicles
Pedestrian & traffic sign detection, lane tracking



Internet Services
Image classification, speech recognition, NLP

Media & Entertainment
Digital content creation

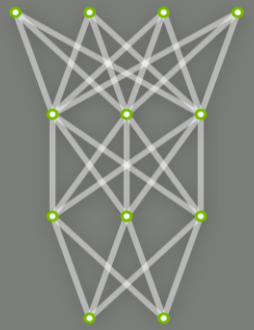


Intelligent Video Analytics
AI cities, urban safety

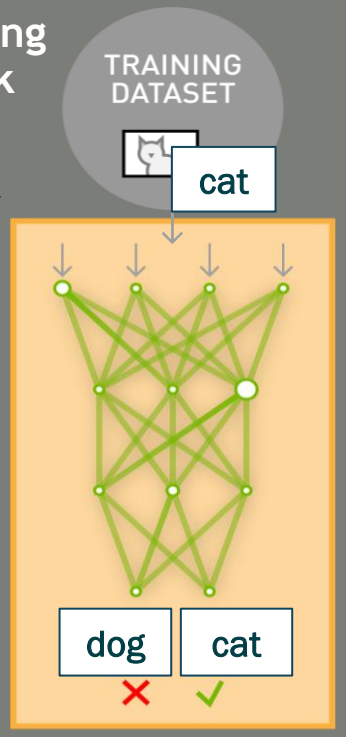
How Does Deep Learning Work?

TRAINING
Learning a new capability from existing data

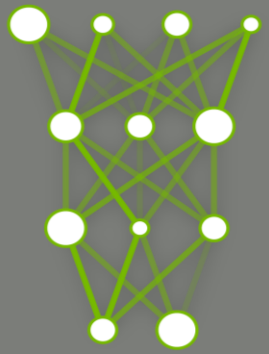
Untrained neural network model



Deep learning framework

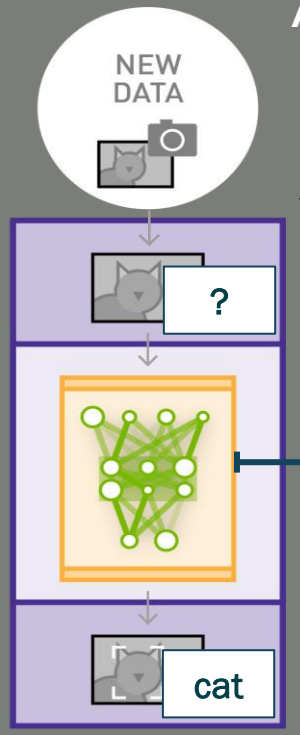


Trained model
New capability



INFERENCE
Applying this capability to new data

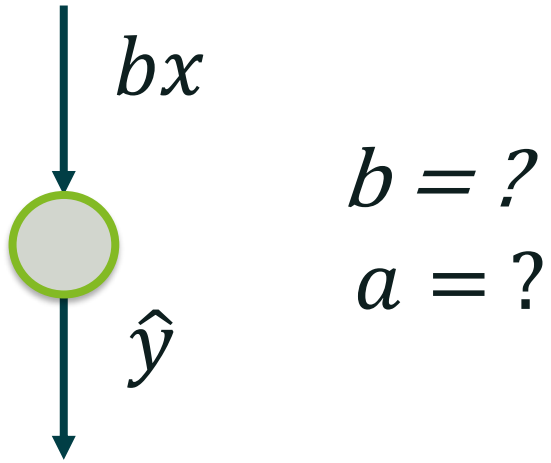
App or service
Featuring capability



Trained model
Optimized for performance

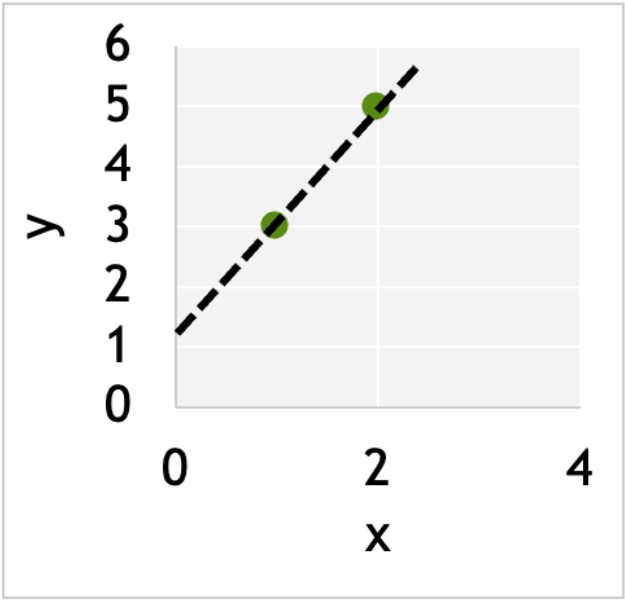
A Simpler Case: Linear Model

$$\hat{y} = bx + a$$



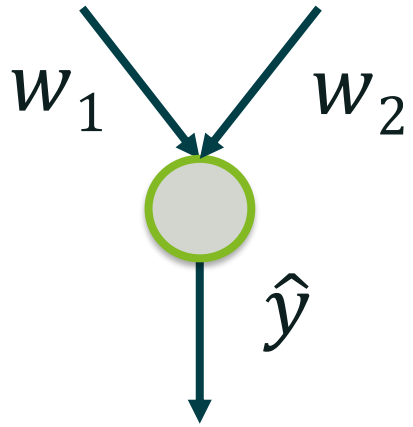
$b = ?$

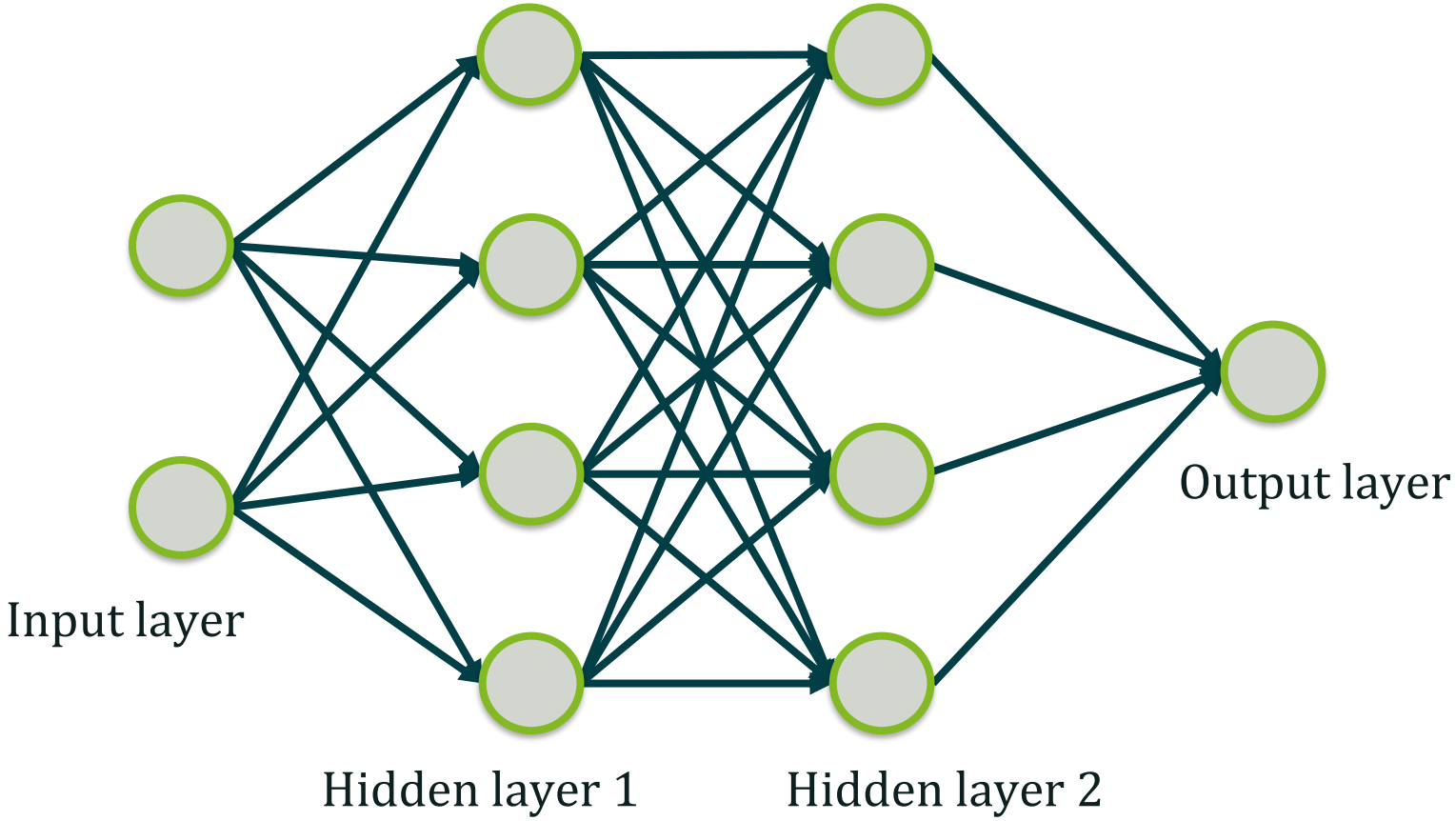
$a = ?$



Weights and ***biases*** are the learnable parameters of a machine learning model

From Neuron to Network

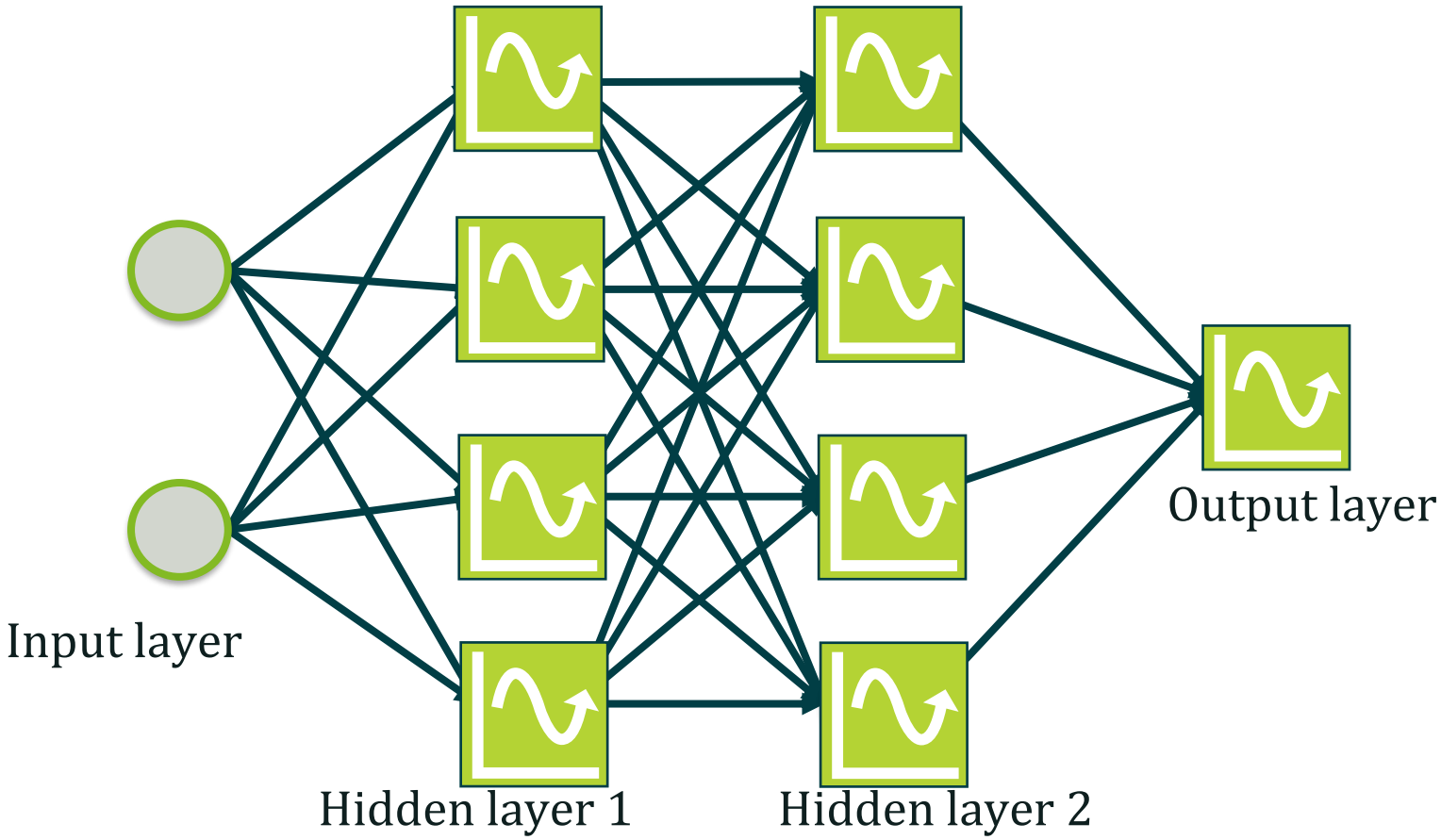




Activation Function

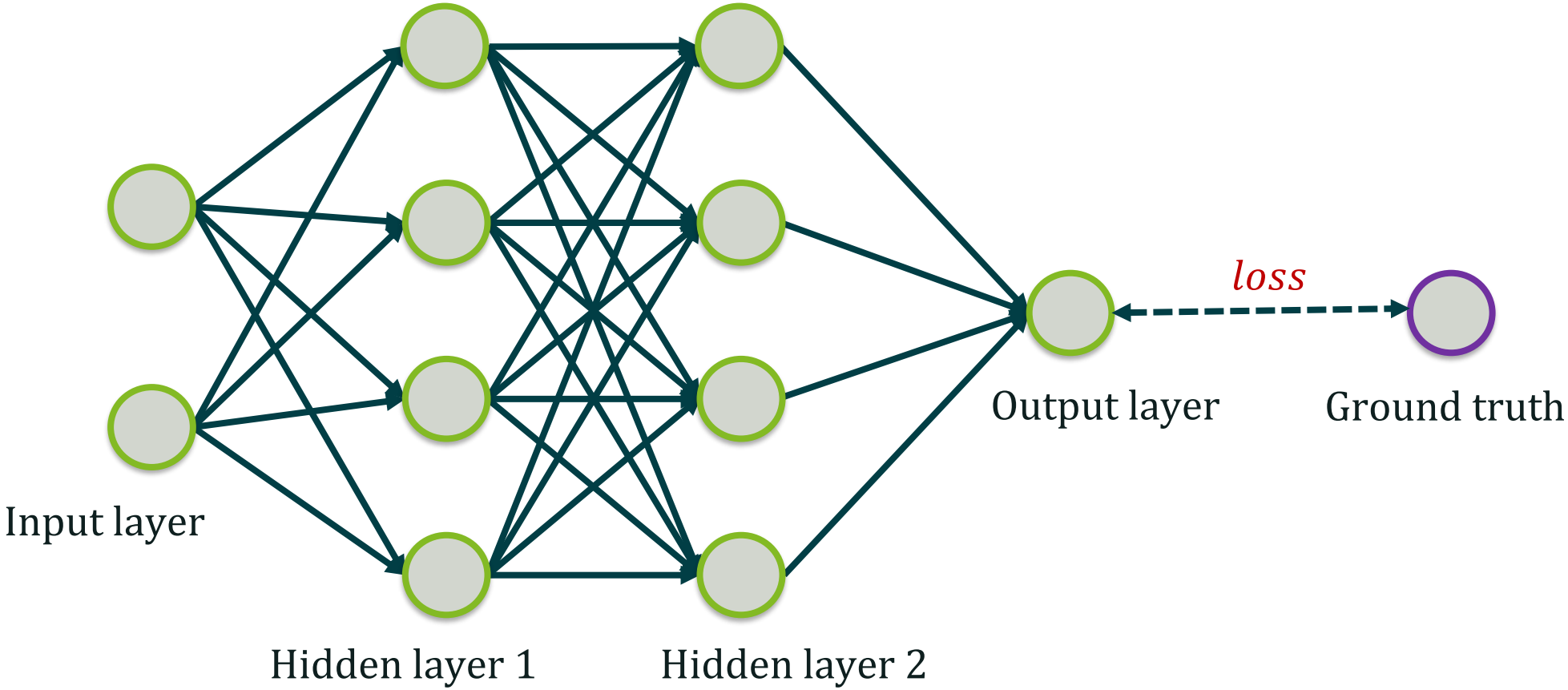
Popular activation functions

- ReLU
- Tanh
- Sigmoid
- …

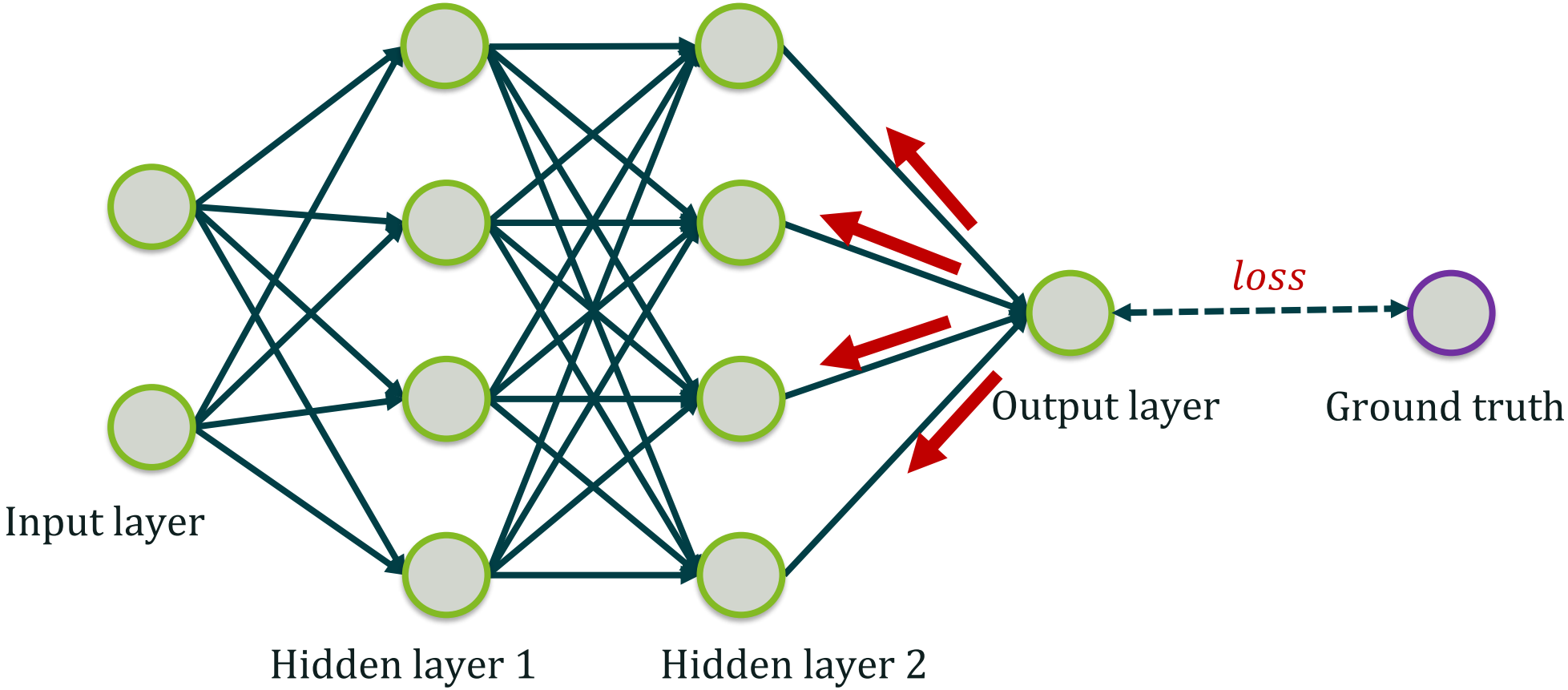


How Does a Neural Network Learn?

Loss

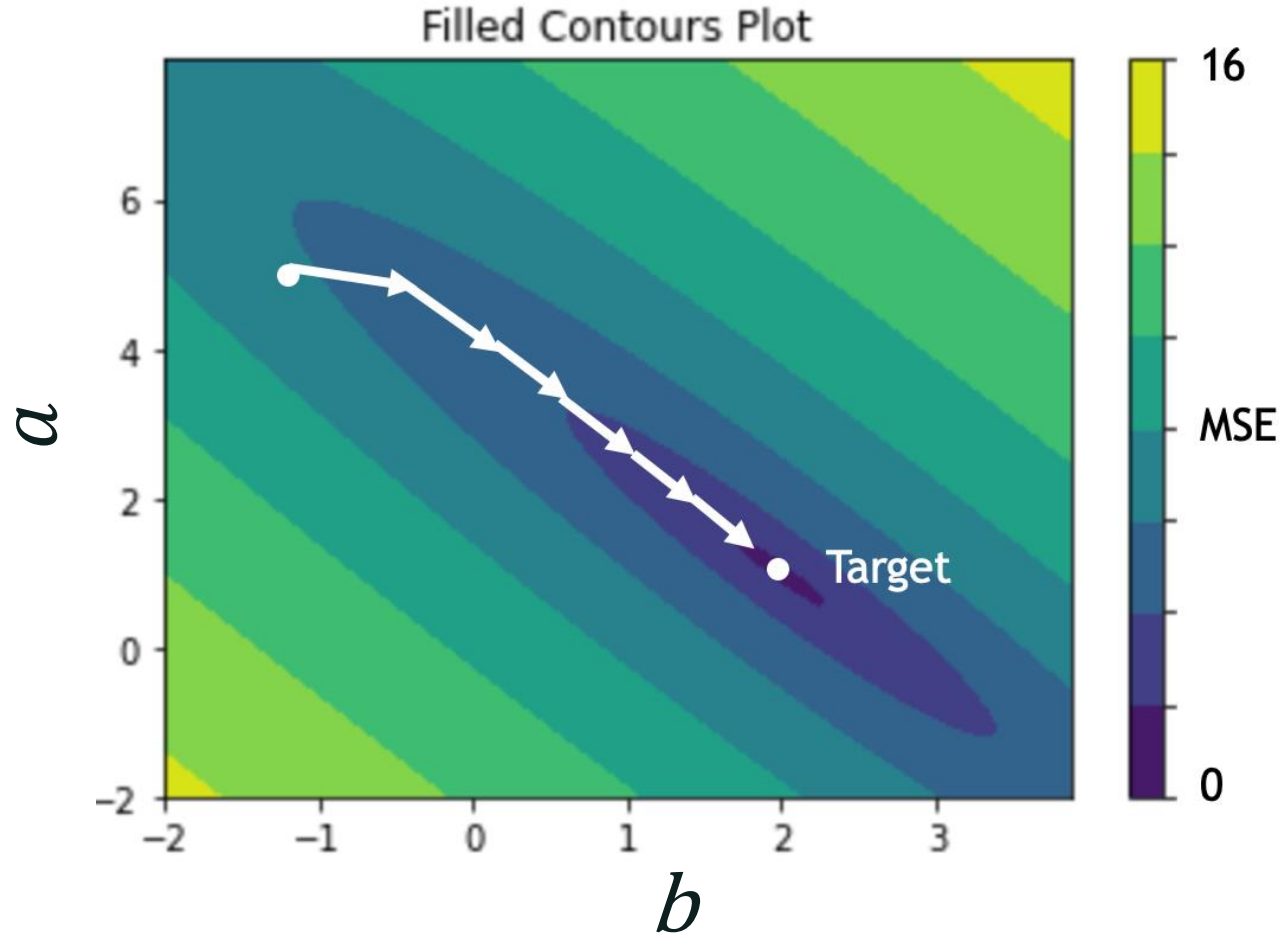


Backpropagation



Gradient Descent

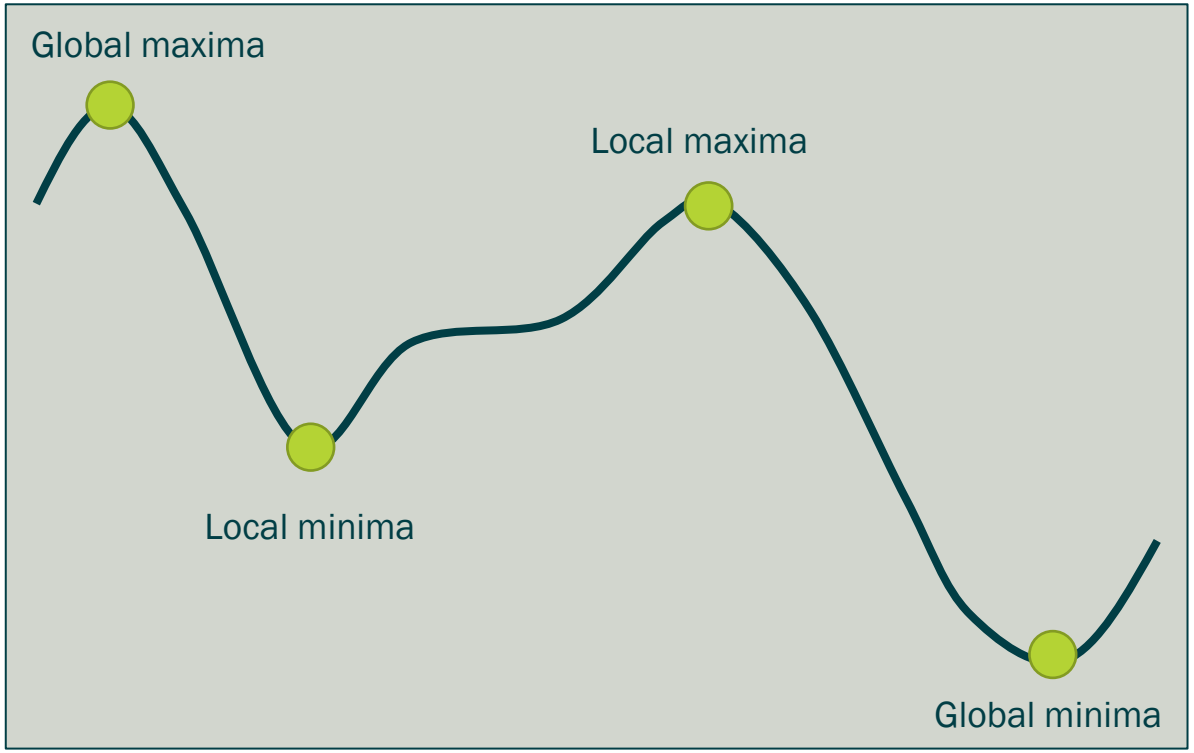
- **Gradient:** which direction loss decreases the most
- **Learning rate:** how far to travel
- **Step:** an update to the weight parameters
- **Epoch:** a model update with the full dataset
- **Batch:** a sample of the full dataset
- **Momentum:** accelerates the optimization process



Optimizers

Some popular optimizers

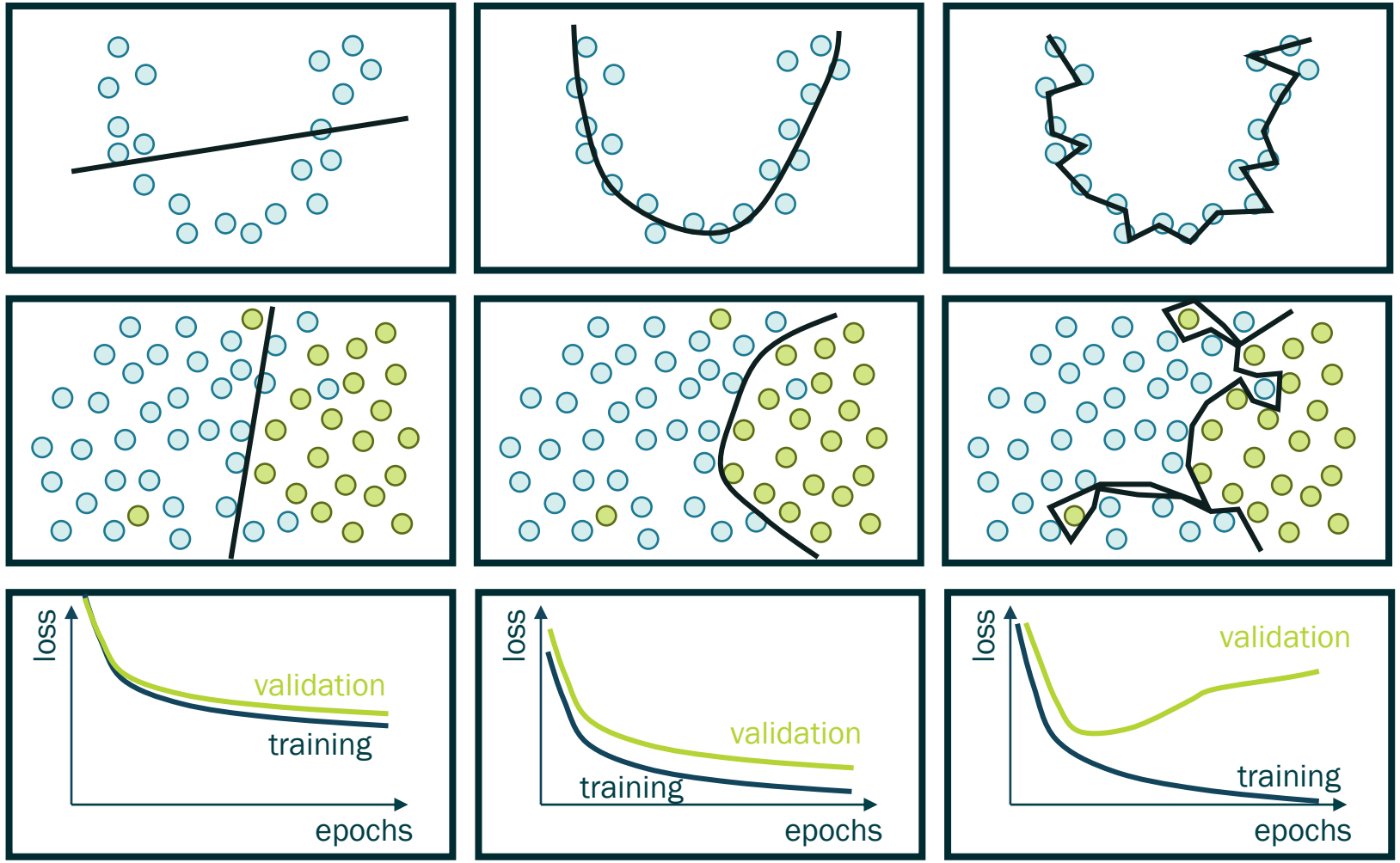
- Adam
- Adagrad
- RMSprop
- Stochastic gradient descent (SGD)
- Madgrad



What Can Go Wrong and How to Make it Right?

NVIDIA

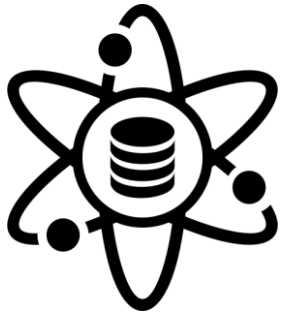
Underfitting – Just Right – Overfitting



Preventing Over- and Underfitting

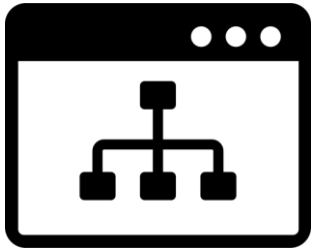
With the help of data

- Add more examples
- Check that your data set is balanced
- Use a separate test set
- Apply data augmentation



Using some techniques

- Early stopping / increasing number of epochs
- Changing network complexity
- Regularization
  - Pruning
  - Dropout
  - Loss penalty (L1 and L2)
- Ensembling
- Transfer learning

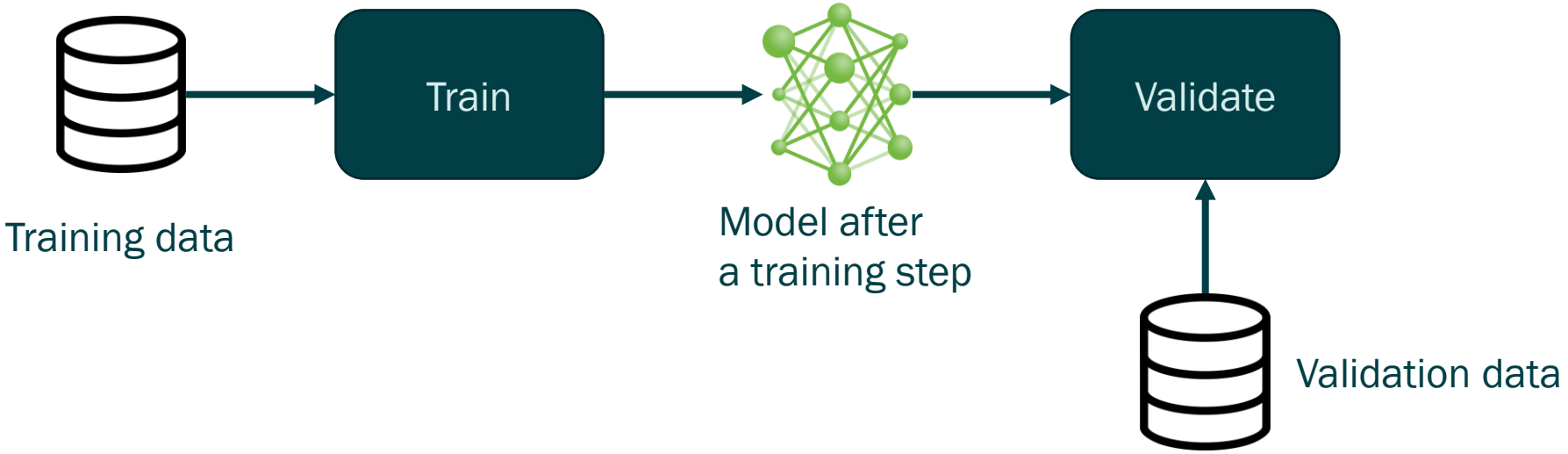


It’s All About Data

NVIDIA

Preparing Datasets for Deep Learning Training

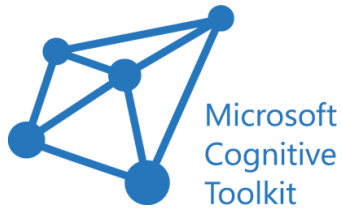
- Acquiring *enough* data
- Ground truth labelling
- Balancing classes
- Splitting data into ***training***, ***validation*** and ***testing*** sets



What's Next?

NVIDIA

Deep Learning Frameworks: Your Starting Point



Summary: In This Talk You Have Learned

- What is a neural network and what does it mean to train it?
- Important steps of the deep learning training process and parameters affecting it
- What can go wrong and ways to make it right
- How to prepare your data for training
- What public resources you can use to master deep learning

Resources

Community resources

Kaggle

https://www.kaggle.com/

Papers with code

https://paperswithcode.com/

MLPerf

https://mlperf.org/

From the industry

Google AI Education

https://ai.google/education/

NVIDIA Devblog

https://developer.nvidia.com/blog/

Machine learning course by Stanford (Coursera)

https://www.coursera.org/learn/machine-learning

AI Conferences

NeurIPS

https://nips.cc/

CVF conferences

https://openaccess.thecvf.com/menu

ICML

https://icml.cc/

Thank you!

NVIDIA