Multi-modal data integration using deep learning and applications in precision oncology

Dr. Bora Uyar

Computational Genomics Workshop

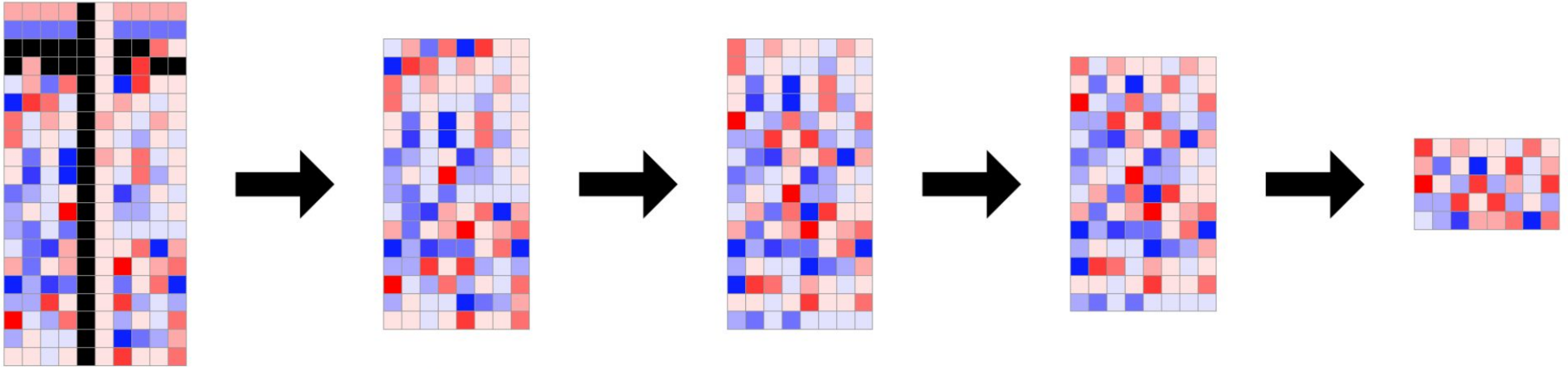
10-16 March 2025

Bioinformatics & Omics Data Science Platform

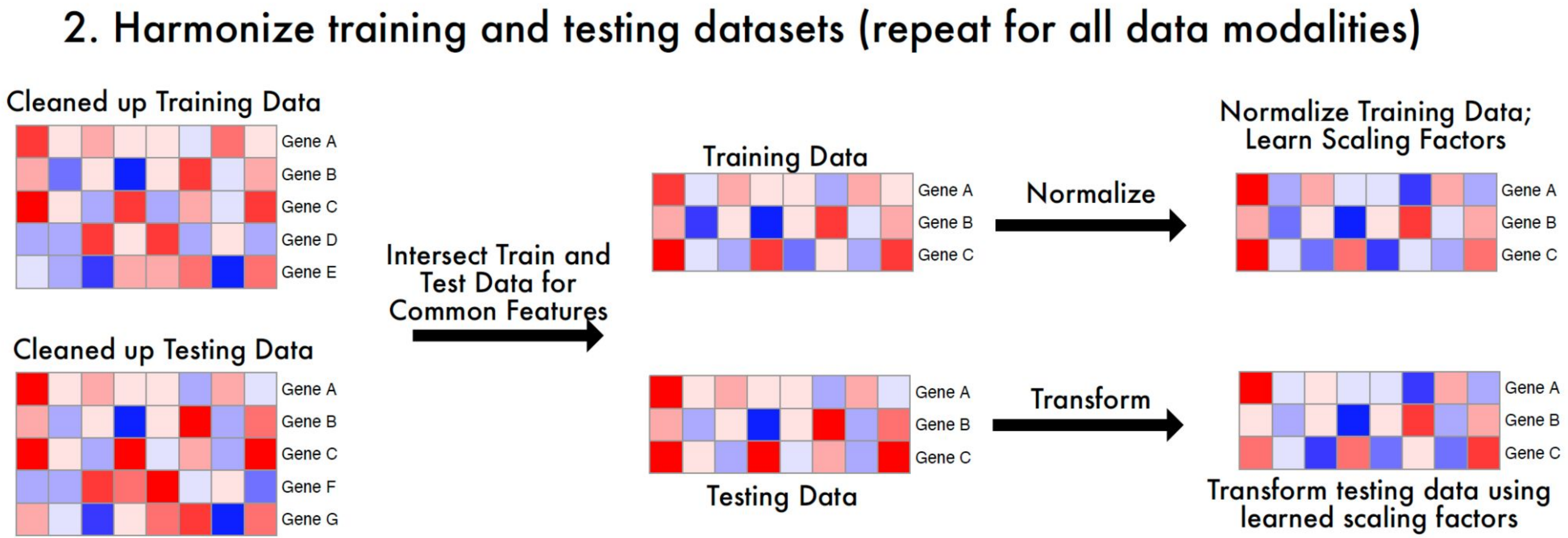
MAX DELBRÜCK CENTER

Data Import

1. Data import and cleanup (repeat for all data modalities)



Data Import



2. Harmonize training and testing datasets (repeat for all data modalities)
Cleaned up Training Data
Gene A
Gene B
Gene C
Gene D
Gene E
Cleaned up Testing Data
Gene A
Gene B
Gene C
Gene F
Gene G
Intersect Train and Test Data for Common Features
Training Data
Gene A
Gene B
Gene C
Testing Data
Gene A
Gene B
Gene C
Normalize
Transform
Normalize Training Data; Learn Scaling Factors
Gene A
Gene B
Gene C
Gene A
Gene B
Gene C
Transform testing data using learned scaling factors

Important Terms

Neurons/Nodes

Layers: input layer, hidden layers, output layer

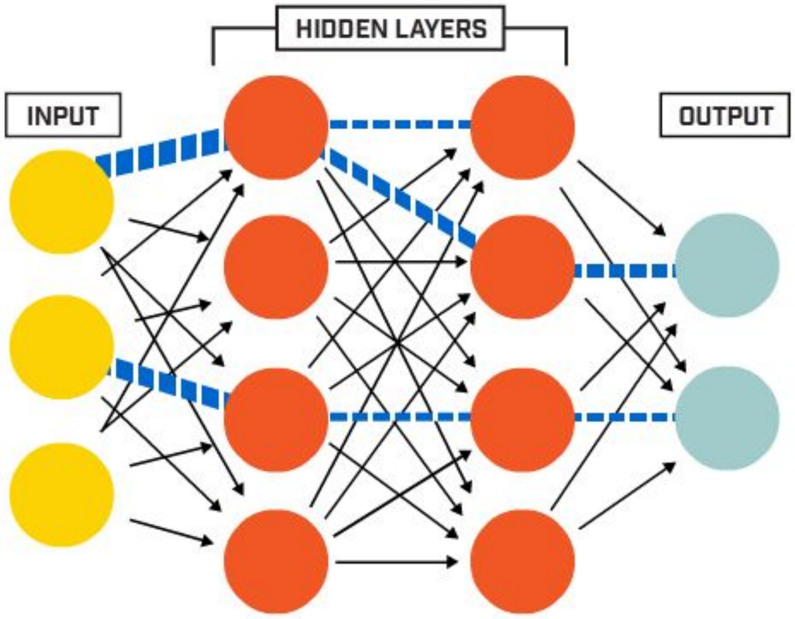
Loss functions

Learning rate

Batch size, batch iterations, epochs

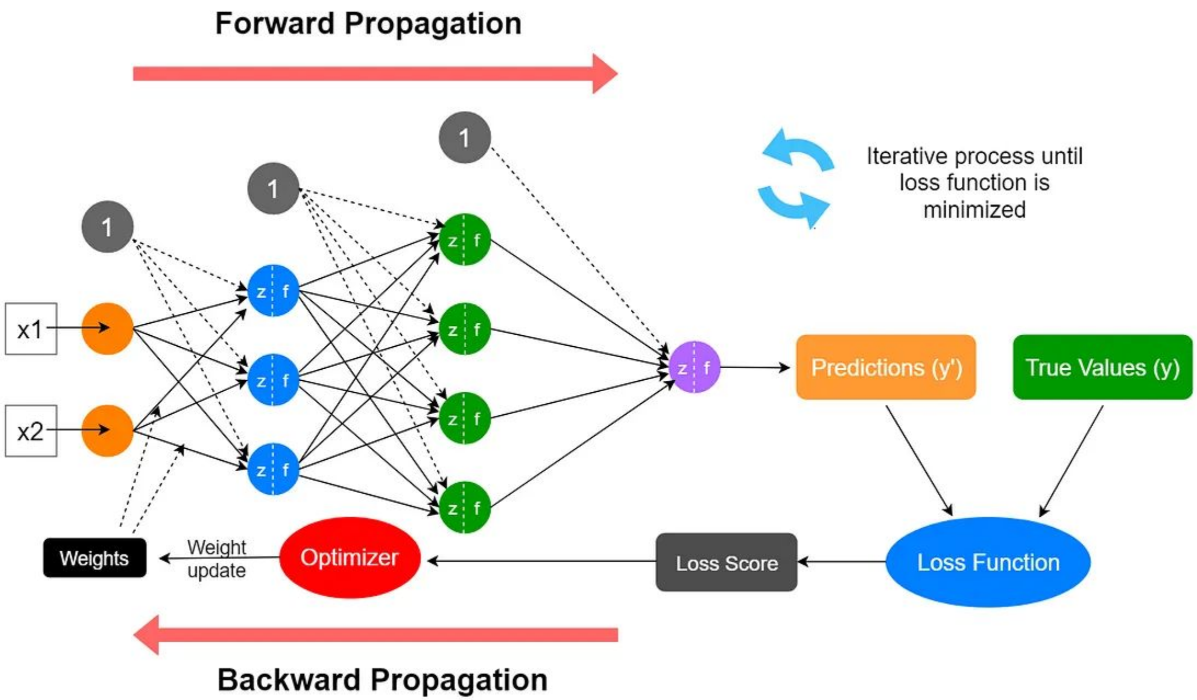
Hyperparameter optimization

Components of a neural network



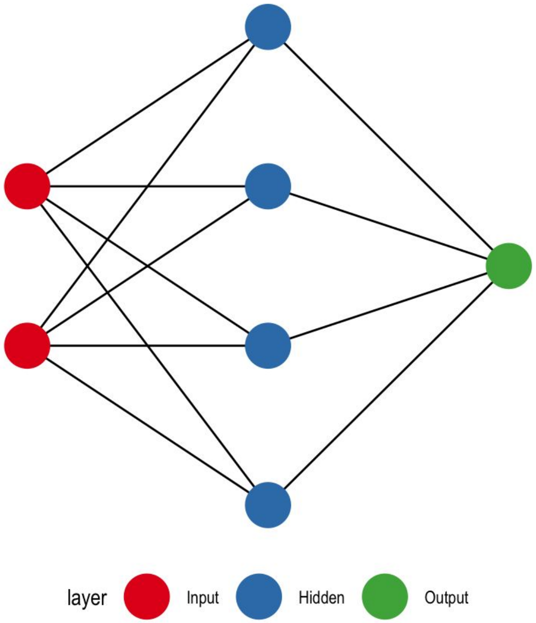
[Introduction to Neural Networks and Their Key Elements (Part-C) — Activation Functions & Layers | Towards AI]()

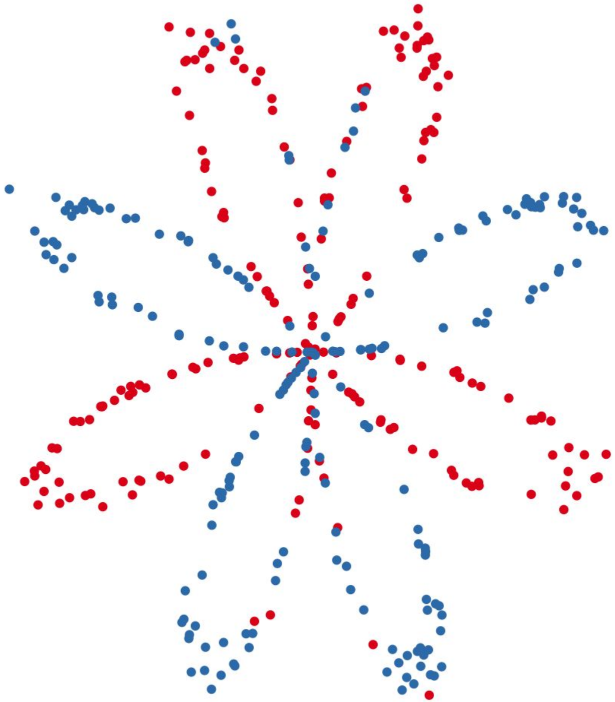
Training Process



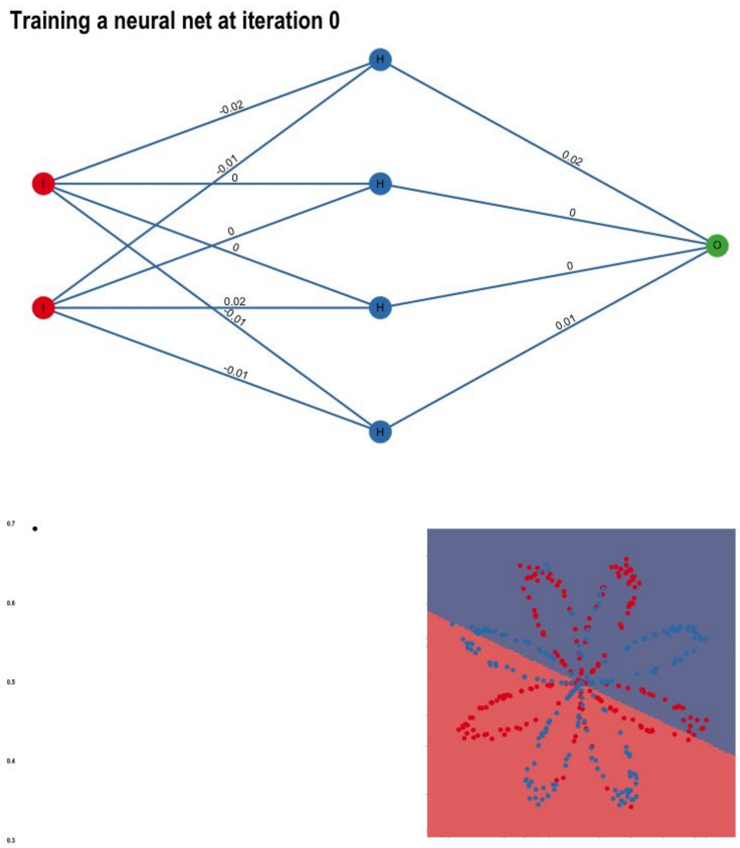
[Decoding Backpropagation and Its Role in Neural Network Learning | ml-articles – Weights & Biases]

Network architecture and the problem



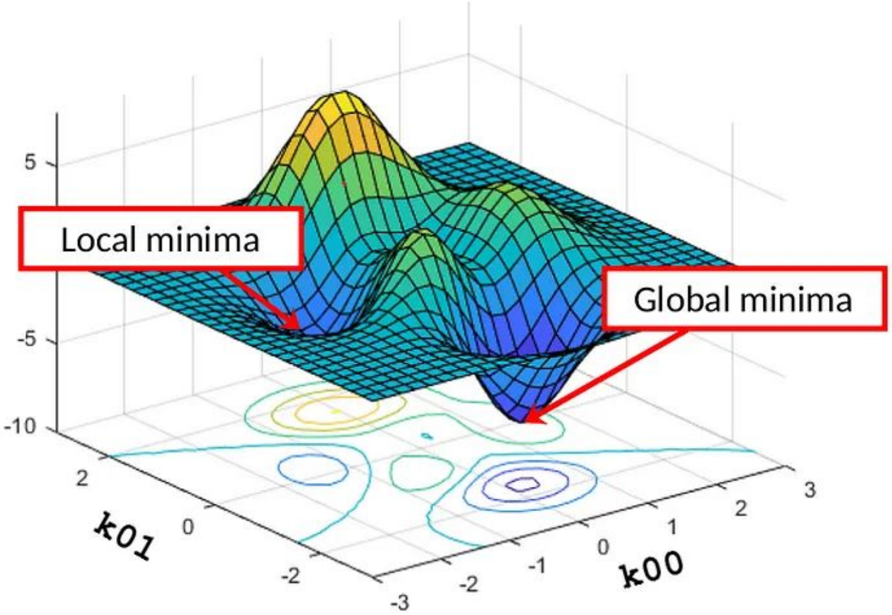


[An animated neuRal net implementation]



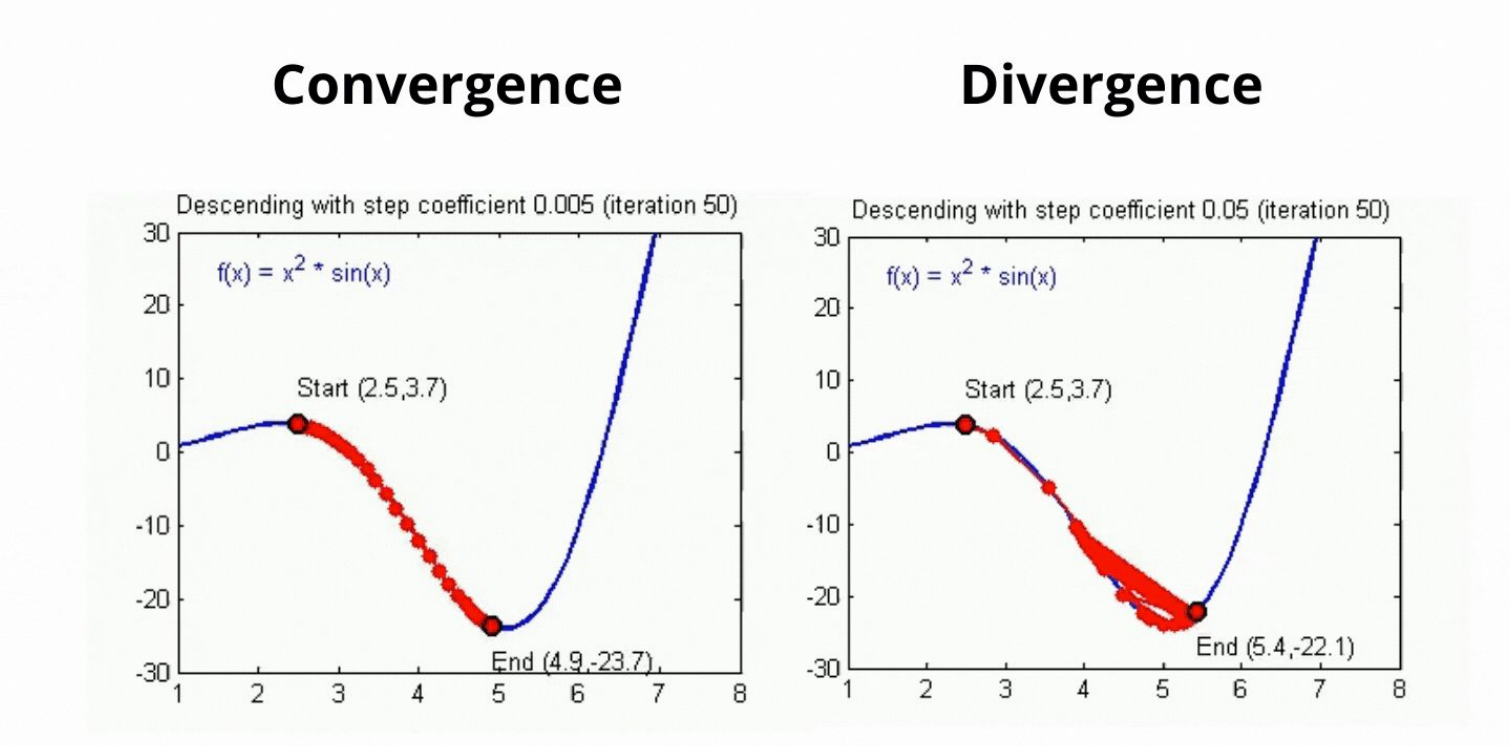
[An animated neuRal net implementation]

Gradient Descent

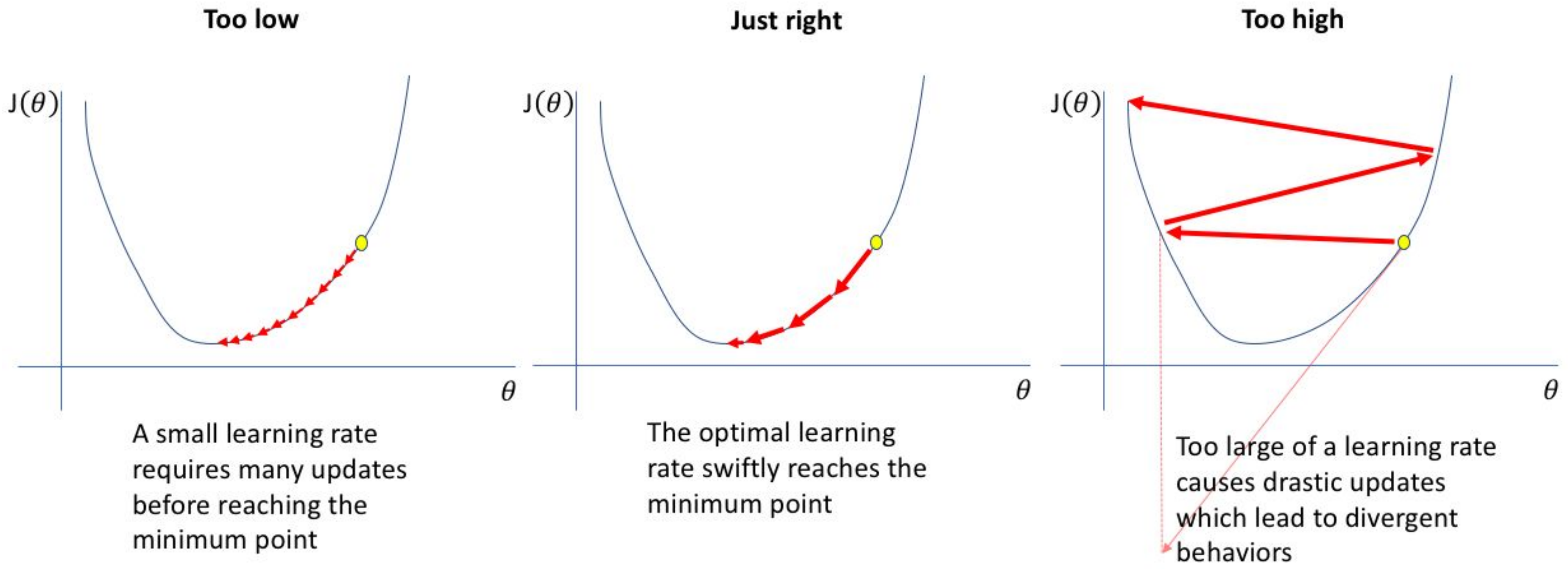


[Deep Learning from Scratch in Modern C++: Gradient Descent | by Luiz doleron | Towards AI]

Impact of the learning rate

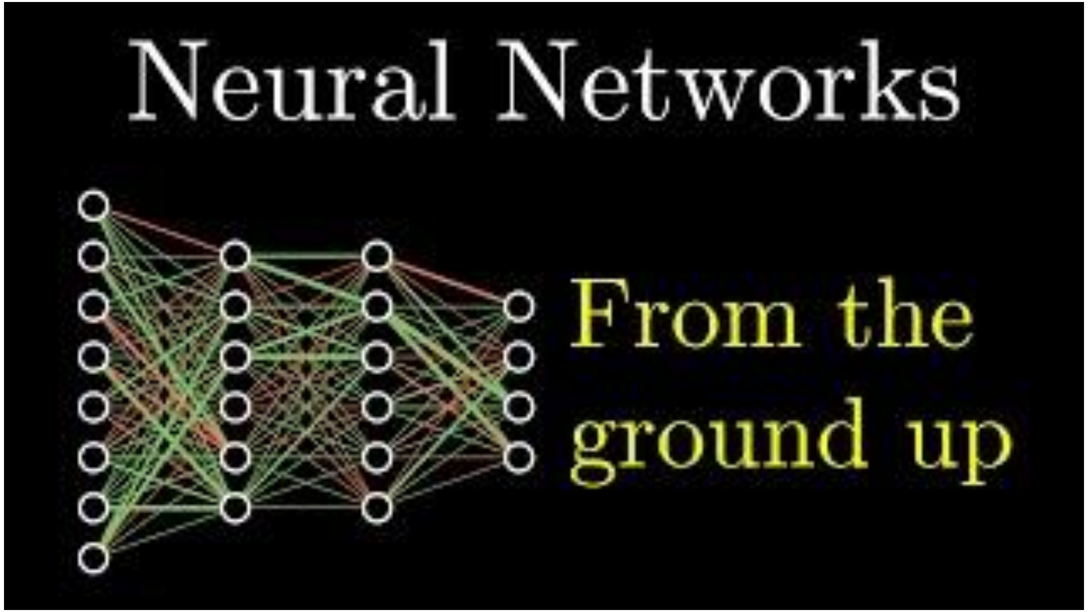


[Mastering Gradient Descent: Optimizing Neural Networks with Precision. | by om pramod | Medium]



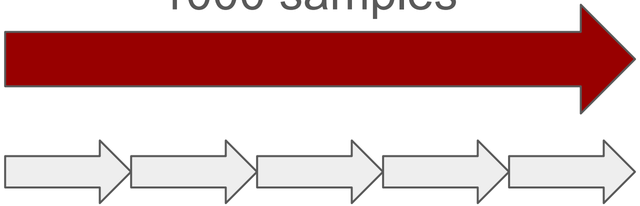
[Setting the learning rate of your neural network.]()

Recommended:



Batch iterations, epochs, HPO iterations

1000 samples



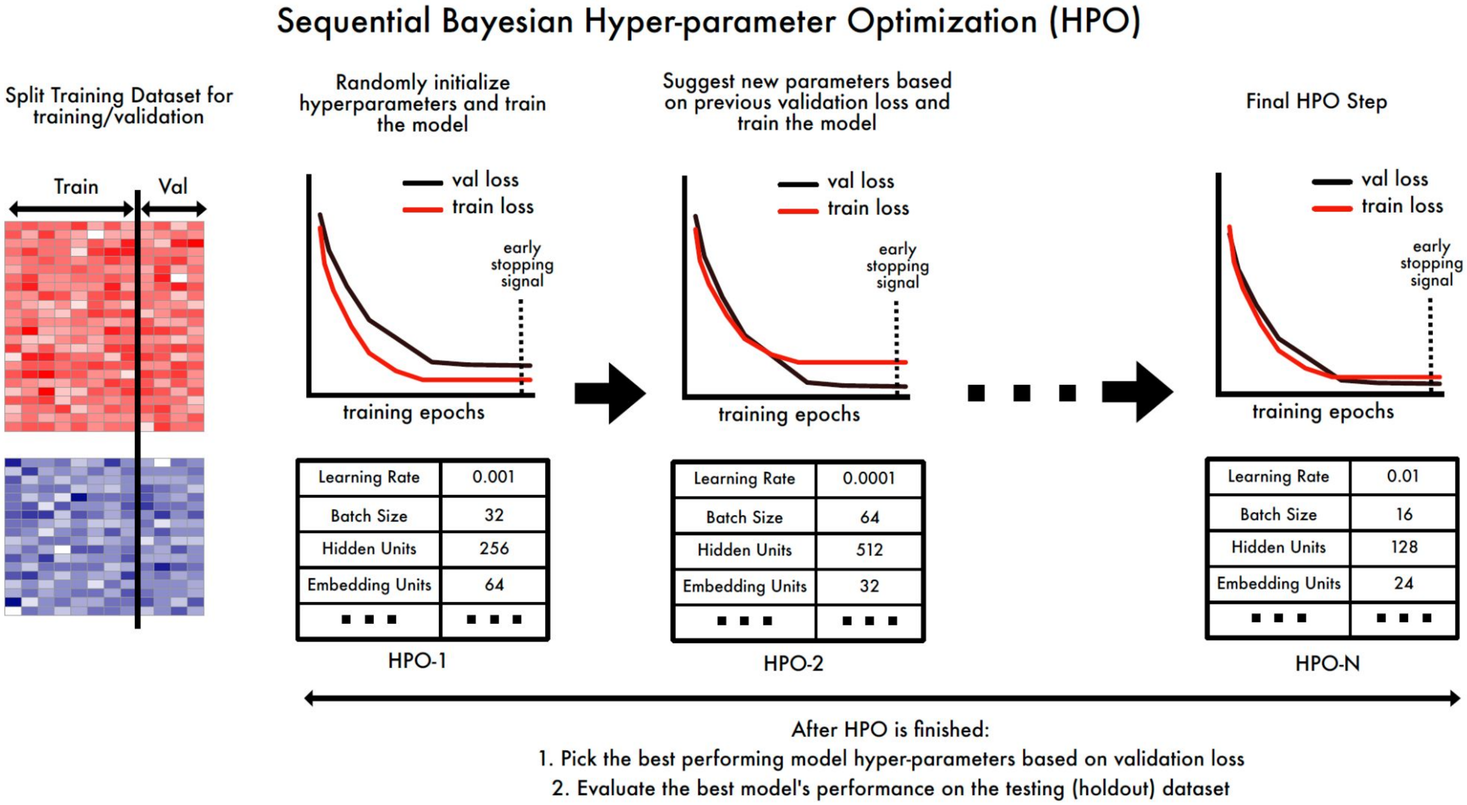
batch size: 200

5 batch iterations = 1 Epoch

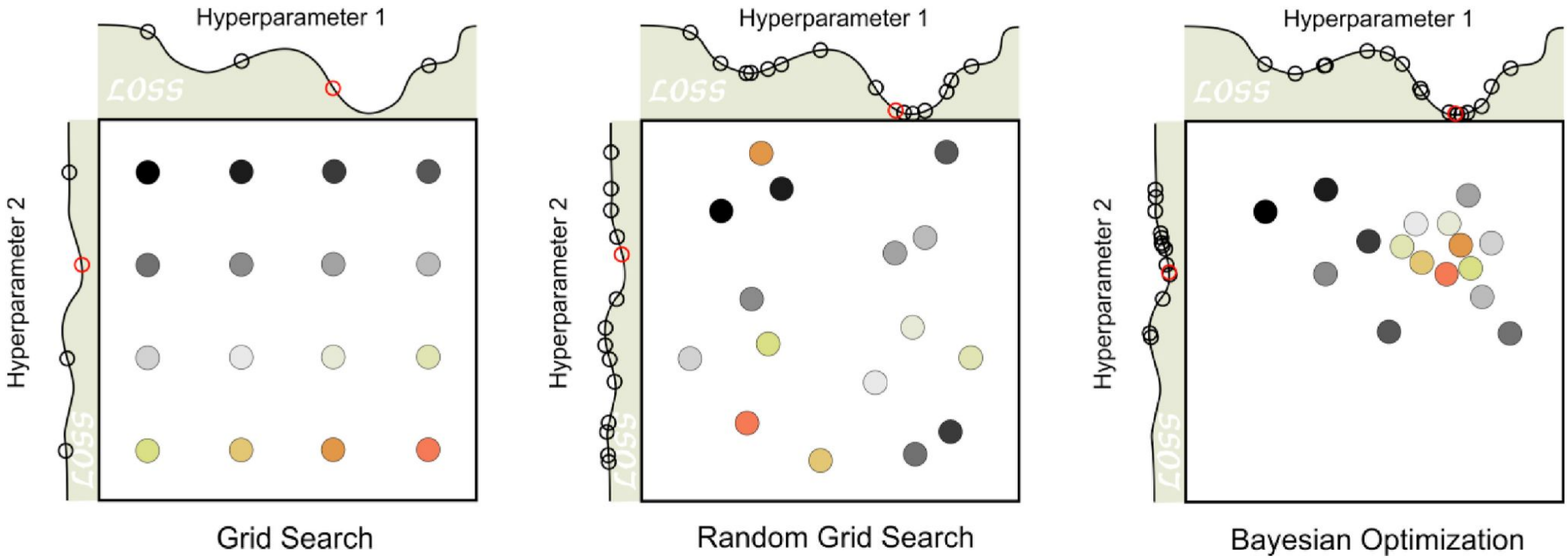
Batch iteration < Epoch < HPO iteration

Epochs, Batch Size, Iterations - How are They Important to Training AI and Deep Learning Models

Hyperparameter Tuning

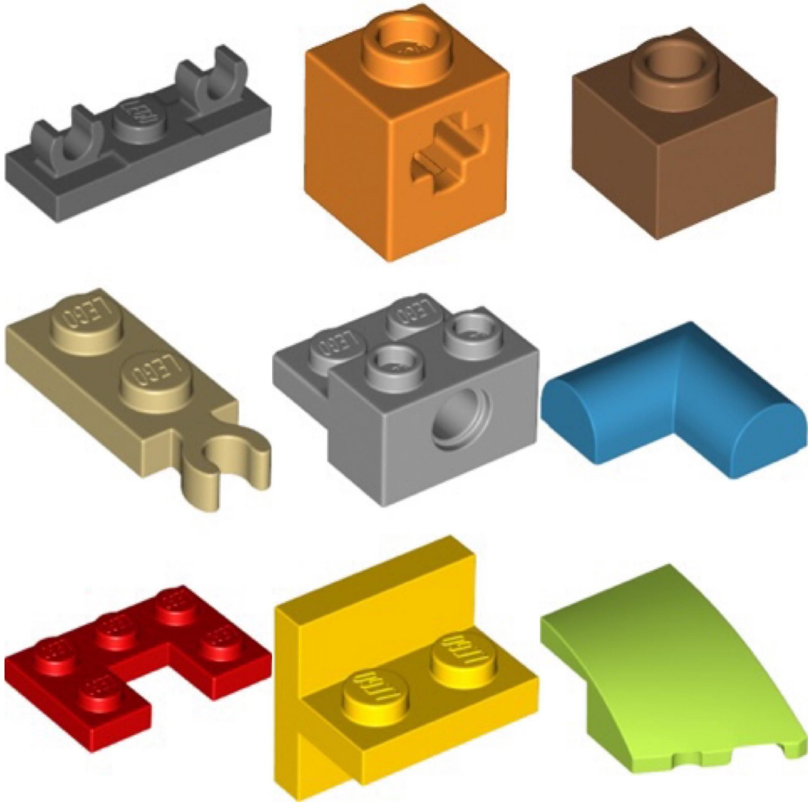


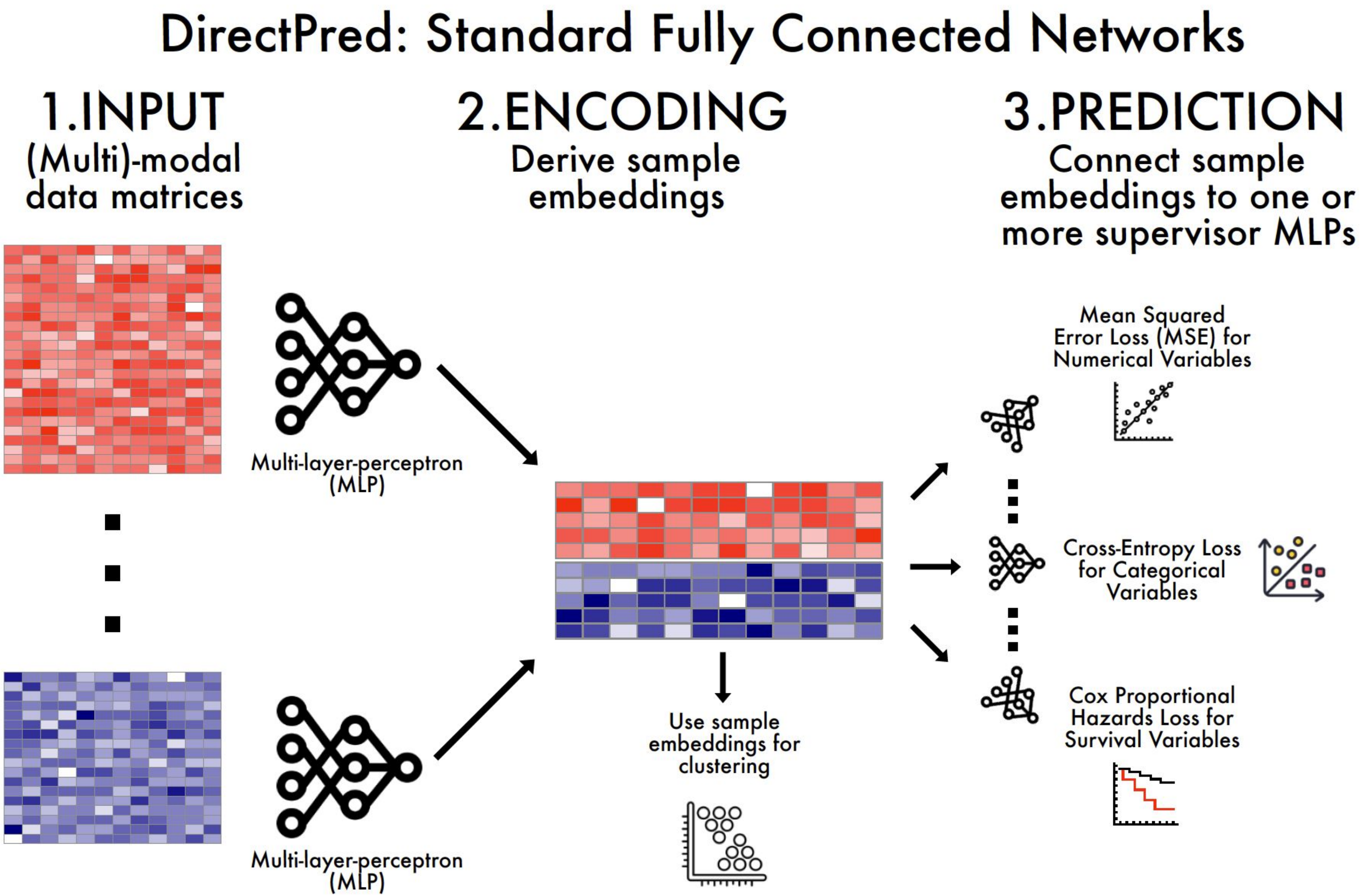
Hyperparameter tuning



[A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks - ScienceDirect]

Neural Network Layers can be combined like Lego pieces




DirectPred: Standard Fully Connected Networks
1.INPUT
(Multi)-modal data matrices
2.ENCODING
Derive sample embeddings
3.PREDICTION
Connect sample embeddings to one or more supervisor MLPs
Multi-layer-perceptron (MLP)
Multi-layer-perceptron (MLP)
Use sample embeddings for clustering
Mean Squared Error Loss (MSE) for Numerical Variables
Cross-Entropy Loss for Categorical Variables
Cox Proportional Hazards Loss for Survival Variables

Homework

https://github.com/BIMSBbioinfo/compgen_course_2025_module3/tree/main/homeworks/hw2