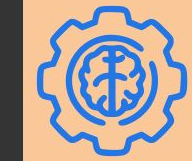
welcome to ML Study Jam

session #4

Google Developer Student Clubs
Simon Fraser University



Timeline

20 June	27 June	4 July	11 July
Session 1	**Session 2**	**Session 3**	**Session 4**
Intro to Machine Learning	Feature Engineering	ML Application	Practice Project
Data Processing	Intro To Deep Learning	Computer Vision	

Review

Model Hypertuning

Review

Model Hypertuning

Data Preprocessing

Importance of data preprocessing: Enhancing data quality for better model performance.

Handling missing data: Managing and filling gaps in the dataset.

Feature scaling and normalization: Scaling features to a common range for optimal model training.

One-hot encoding and categorical variables: Transforming categorical data into numerical format.

Handling outliers: Addressing extreme values that can impact model behavior.

Train-test split and cross-validation: Evaluating model performance on unseen data for robustness.

Machine Learning (ML)

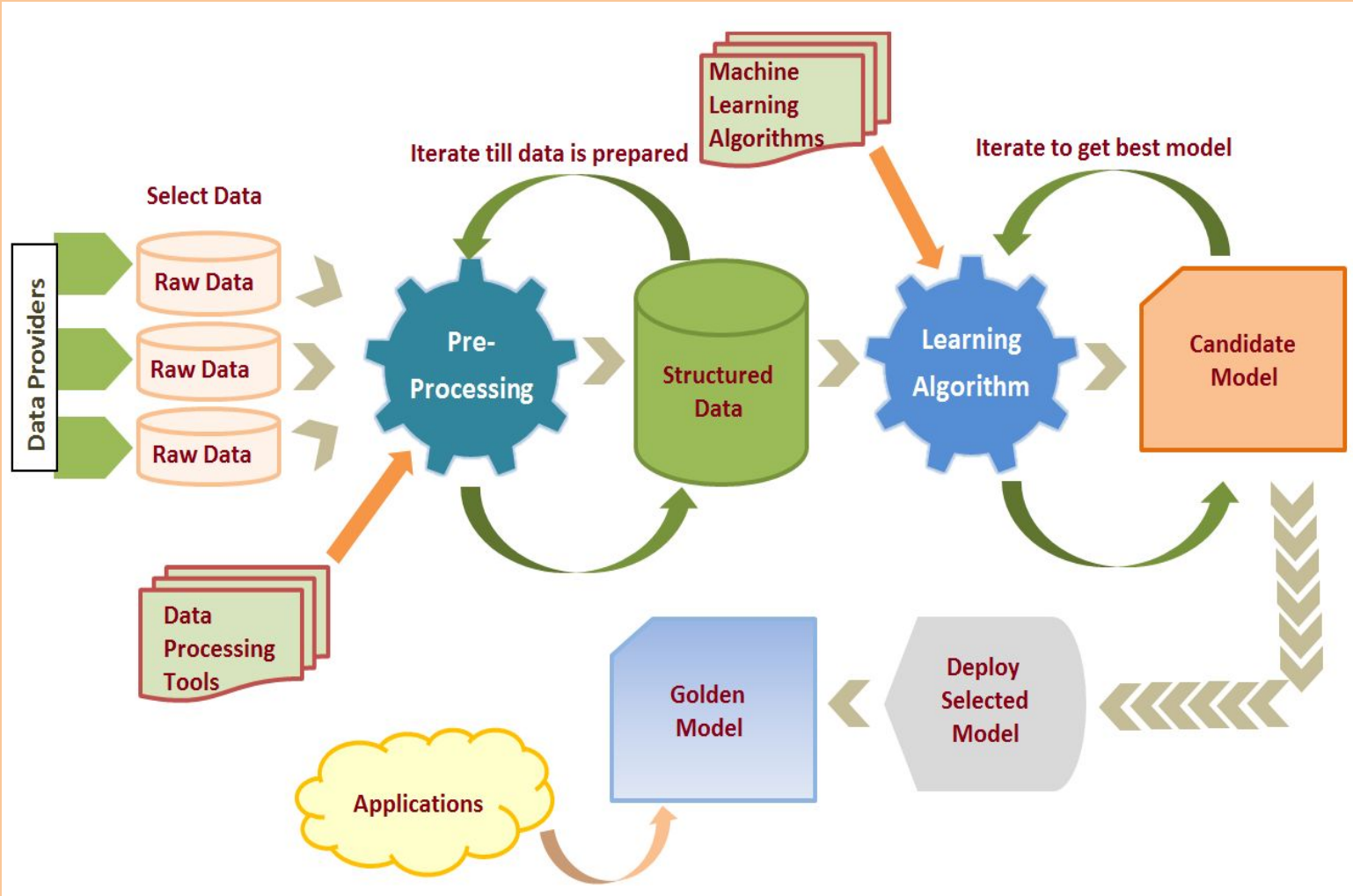
Machine Learning is the field of study that focuses on enabling computers to learn from data and make predictions or decisions without being explicitly programmed.

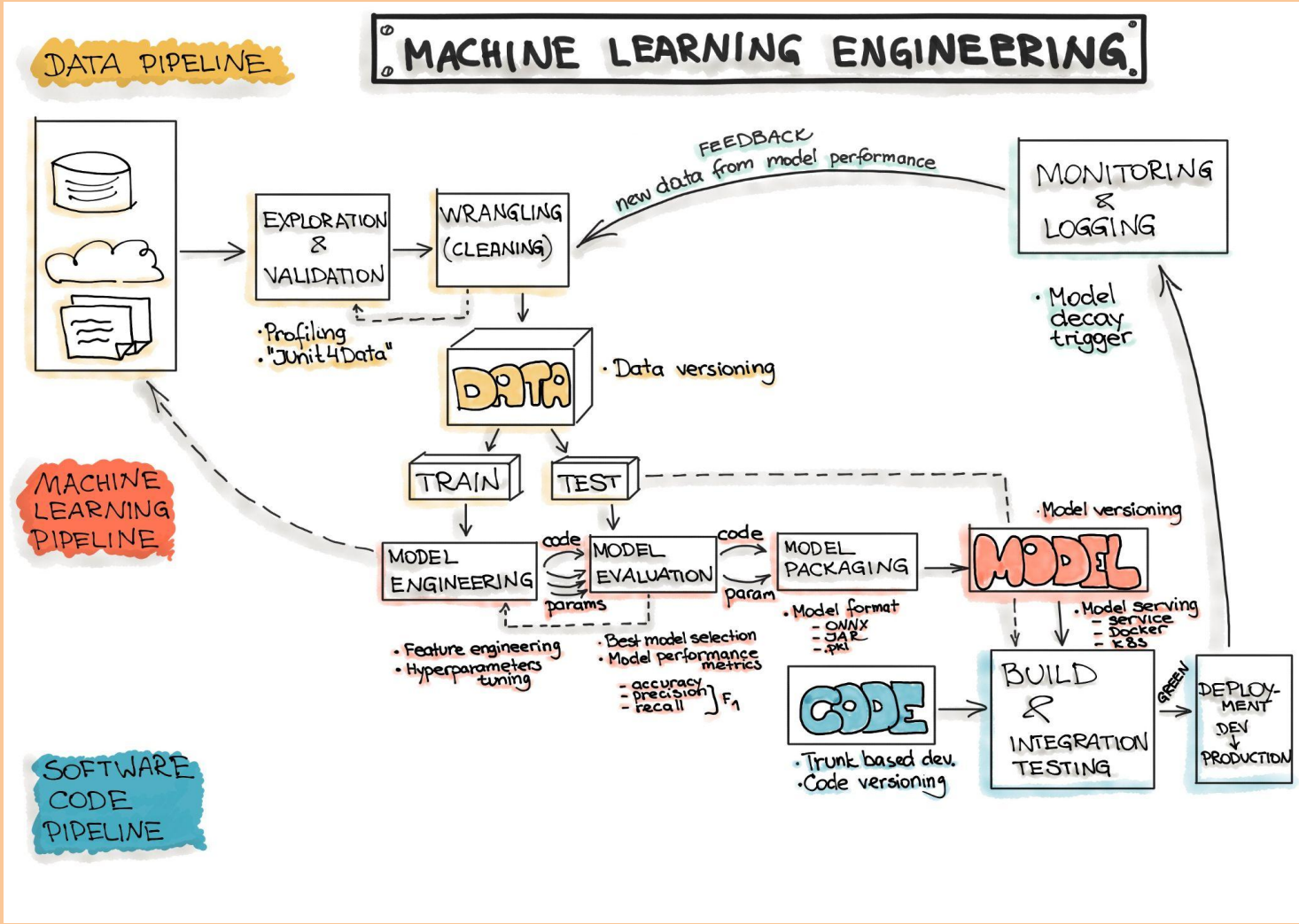
Key Concepts: Supervised Learning vs. Unsupervised Learning

Classification Algorithms: Decision Trees, Random Forests, KNN

Regression Algorithms: Linear Regression, Polynomial Regression

Evaluation Metrics: Accuracy, Precision, Recall, F1 Score


Machine Learning Algorithms
Iterate till data is prepared
Iterate to get best model
Select Data
Data Providers
Raw Data
Raw Data
Raw Data
Pre-Processing
Structured Data
Learning Algorithm
Candidate Model
Data Processing Tools
Golden Model
Deploy Selected Model
Applications


DATA PIPELINE
MACHINE LEARNING ENGINEERING
FEEDBACK
new data from model performance
MONITORING & LOGGING
EXPLORATION & VALIDATION
WRANGLING (CLEANING)
· Profiling
· "Junit4Data"
· Model decay trigger
DATA
· Data versioning
MACHINE LEARNING PIPELINE
TRAIN
TEST
· Model versioning
MODEL ENGINEERING
code
params
MODEL EVALUATION
code
param
MODEL PACKAGING
MODEL
· Model format
- ONNX
- JAR
- .pkl
· Model serving
- service
- Docker
- K8s
· Feature engineering
· Hyperparameters tuning
· Best model selection
· Model performance metrics
- accuracy
- precision
- recall
F1
CODE
· Trunk based dev.
· Code versioning
BUILD & INTEGRATION TESTING
GREEN
DEPLOY-MENT
DEV
PRODUCTION
SOFTWARE CODE PIPELINE

Deep Learning

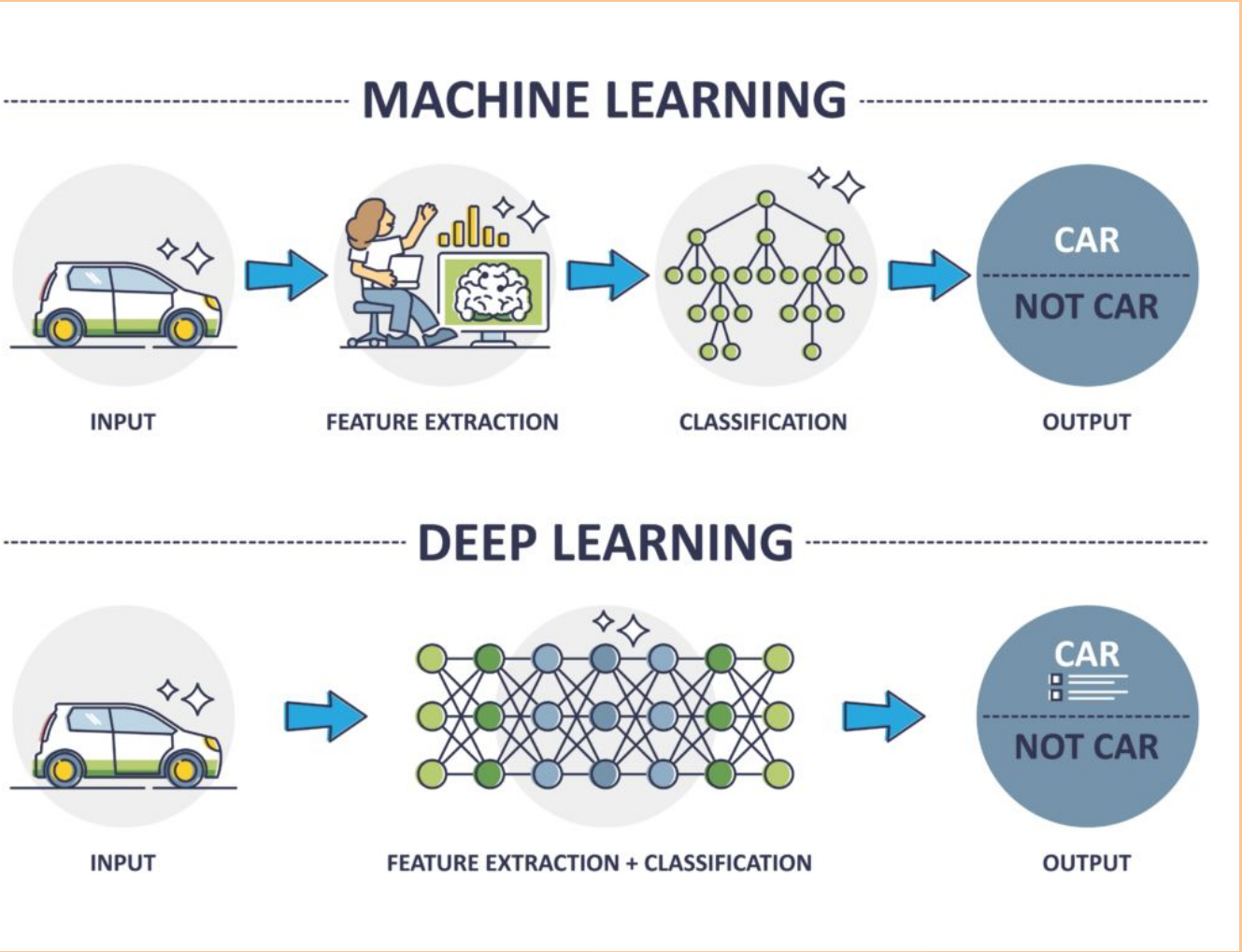
Introduction to Neural Networks: Artificial models inspired by the human brain, composed of interconnected nodes or neurons.

Feedforward Neural Networks: Neural networks where information flows in one direction, from input to output.

Convolutional Neural Networks (CNNs): Specialized neural networks for image processing tasks, leveraging convolutional layers to extract spatial data.

Recurrent Neural Networks (RNNs): Neural networks designed to process sequential data, capturing temporal dependencies.

Transfer Learning and Pretrained Models: Leveraging pre-trained models trained on large datasets to solve related tasks with limited data.


MACHINE LEARNING
CAR
NOT CAR
INPUT
FEATURE EXTRACTION
CLASSIFICATION
OUTPUT
DEEP LEARNING
CAR
NOT CAR
INPUT
FEATURE EXTRACTION + CLASSIFICATION
OUTPUT

Computer Vision

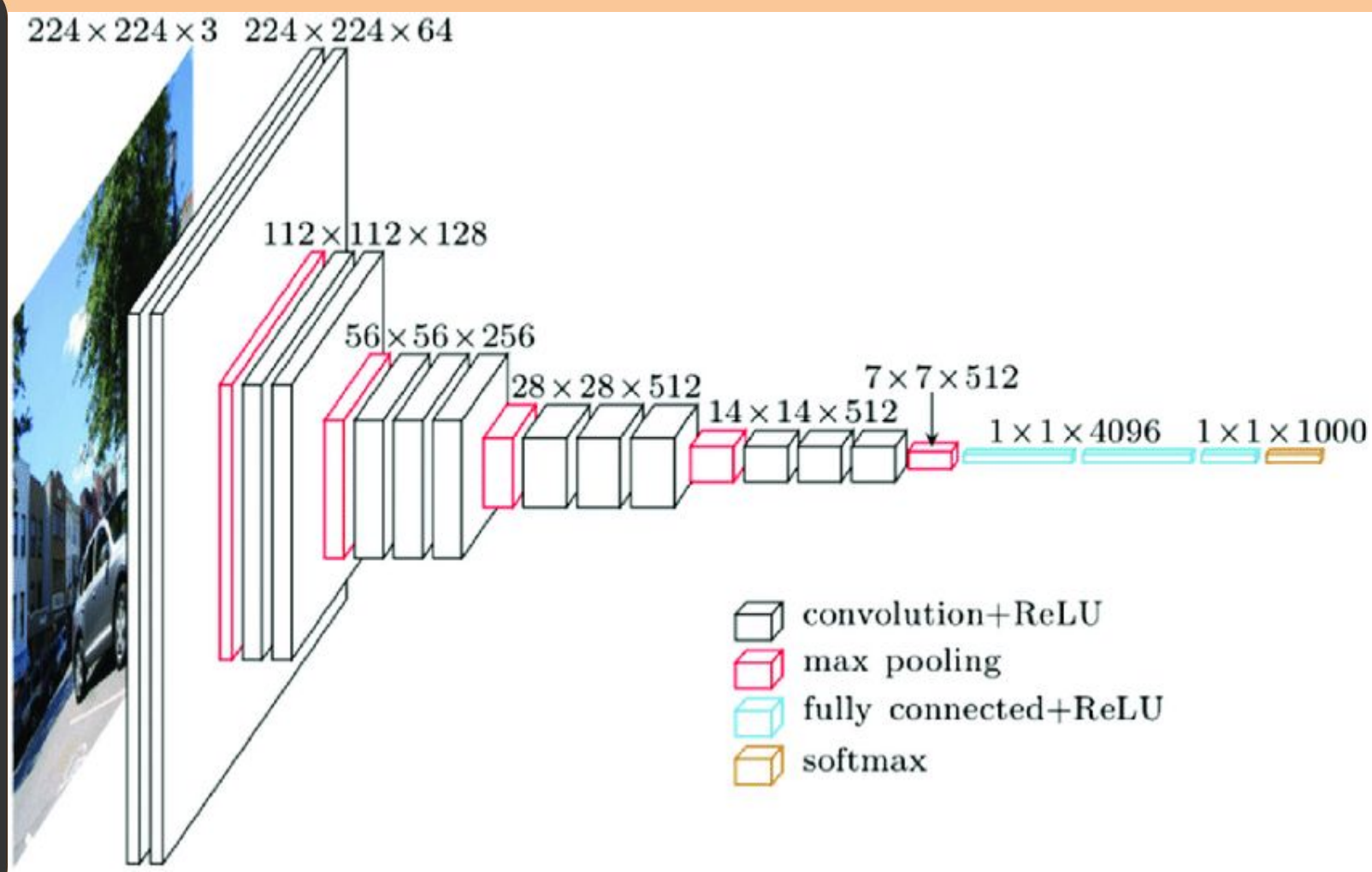
The field of AI focused on teaching computers to interpret and understand visual data.

Applications: Object recognition, image classification, image segmentation, face detection, etc.

Image Preprocessing Techniques: Normalization, resizing, cropping, data augmentation.

Object Detection and Image Segmentation: Techniques to identify and localize objects within images.

CNN architectures designed specifically for image classification tasks, such as AlexNet, VGGNet, and ResNet.


224 × 224 × 3
224 × 224 × 64
112 × 112 × 128
56 × 56 × 256
28 × 28 × 512
14 × 14 × 512
7 × 7 × 512
1 × 1 × 4096
1 × 1 × 1000
convolution+ReLU
max pooling
fully connected+ReLU
softmax

Natural Language Processing

The field of AI focused on enabling computers to understand, interpret, and generate human language.

Applications: machine translation, text summarization, chatbots, etc.

Text Preprocessing Techniques: Tokenization, stemming, lemmatization, stop word removal, etc.

Language Modeling and Text Classification: Techniques for predicting the next word in a sequence or classifying text into predefined categories.

Sequence-to-Sequence Models for NLP tasks: Architectures like Encoder-Decoder models and Transformers, used for machine translation and other sequence generation tasks.

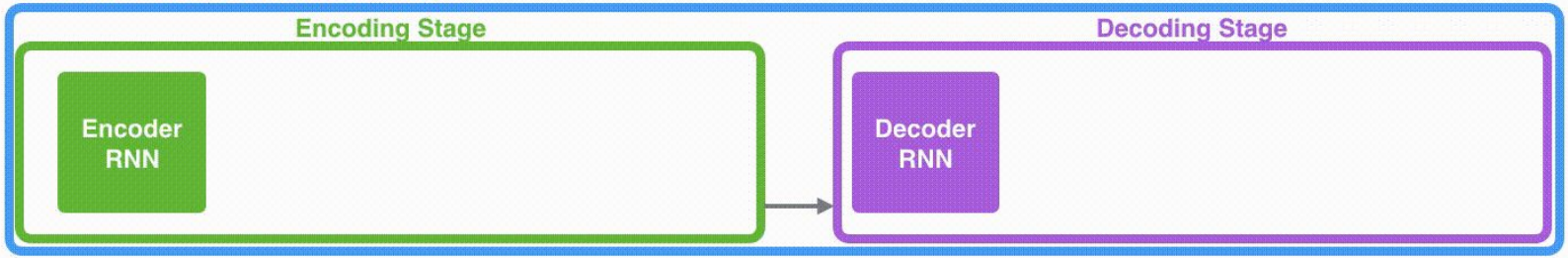
Natural Language Processing cont'd

Language Modeling and Text Classification: Techniques for predicting the next word in a sequence or classifying text into predefined categories.

Sequence-to-Sequence Models for NLP tasks: Architectures like Encoder-Decoder models and Transformers, used for machine translation and other sequence generation tasks.

Neural Machine Translation

SEQUENCE TO SEQUENCE MODEL



Je suis étudiant

Review

Model Hyperparameters

Training Hyperparameters

Learning Rate: The learning rate determines the step size in the optimization process.

Weight Decay: Weight decay is a regularization technique that penalizes large weight values to prevent overfitting.

Batch Size: Batch size represents the number of training examples processed before updating the model's weights.

Number of Epochs: The number of epochs determines how many times the model sees the entire training dataset.

Training Hyperparameters cont'd

Optimizer: The optimizer algorithm updates the model's parameters during training.

Regularization Parameters: Regularization parameters control the complexity of the model to prevent overfitting.

Diminishing Gradient: The problem of diminishing gradients occurs when gradients become extremely small during training.

Exploding Gradient: The problem of exploding gradients occurs when gradients become extremely large.

Exercise Time

https://colab.research.google.com/drive/1O6OEuGmAx1kiHxSmPlwjMd-paOzA47O_?usp=sharing

Kaggle Resources

- https://www.kaggle.com/learn/data-visualization
- https://www.kaggle.com/learn/time-series
- https://www.kaggle.com/learn/machine-learning-explainability

Thank you!

Google Developer Student Clubs
Simon Fraser University