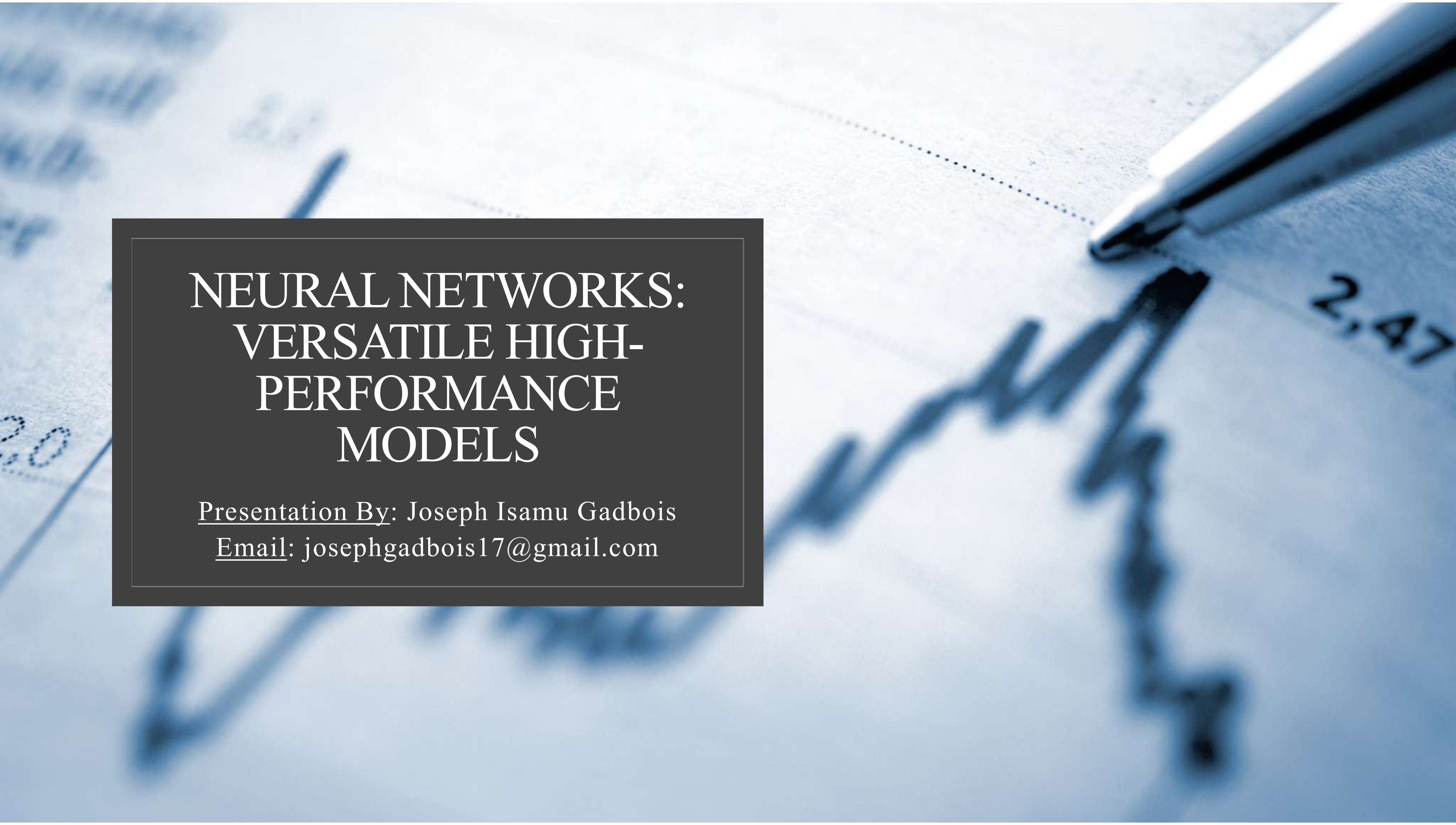
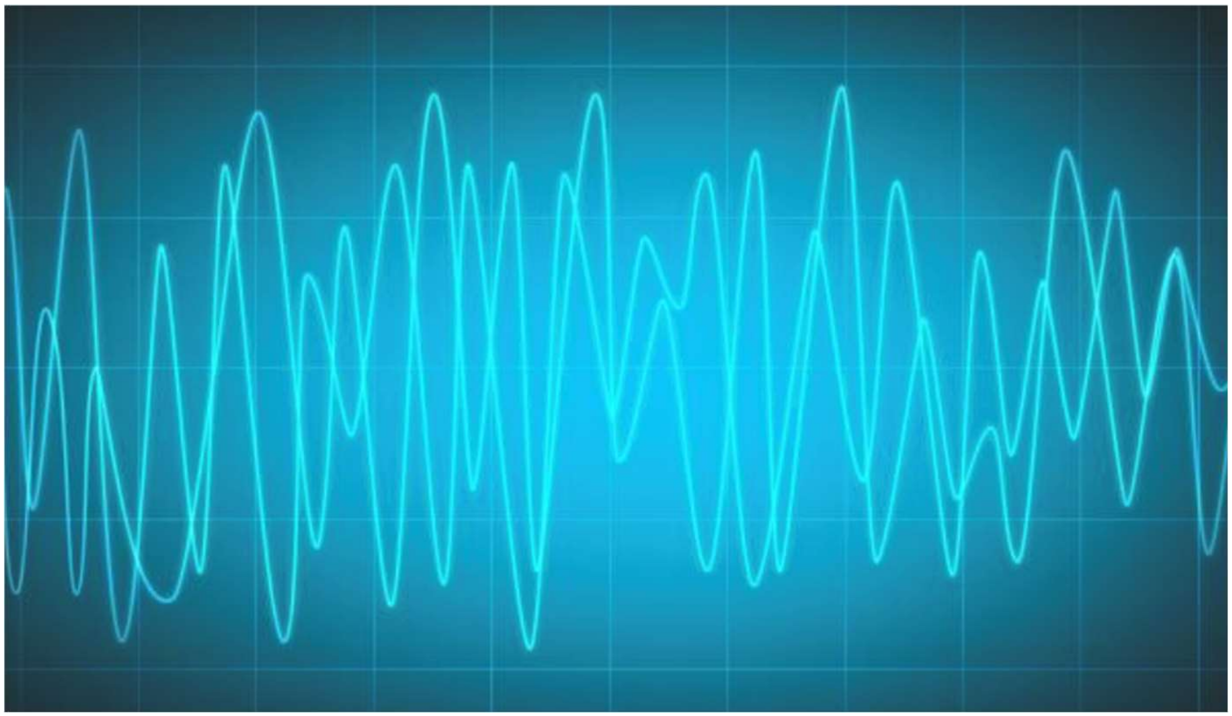


NEURAL NETWORKS: VERSATILE HIGH-PERFORMANCE MODELS

Presentation By: Joseph Isamu Gadbois
Email: josephgadbois17@gmail.com

Digital Signal Data



- **Signal:** sensory data that come from a function that conveys information about the state or behavior of a physical system.
- **1D Signals:** sequential data; time series, stochastic process.
- **2D Signals:** spatial data; images.

Applications and Models

Image Data Applications:

Image Classification

- Convolutional Neural Network
- Linear Discriminant Analysis

Image Segmentation

- U-Net Fully Convolutional Network
- K-Means Clustering

Sequential Data Applications:

Time Series Forecasting

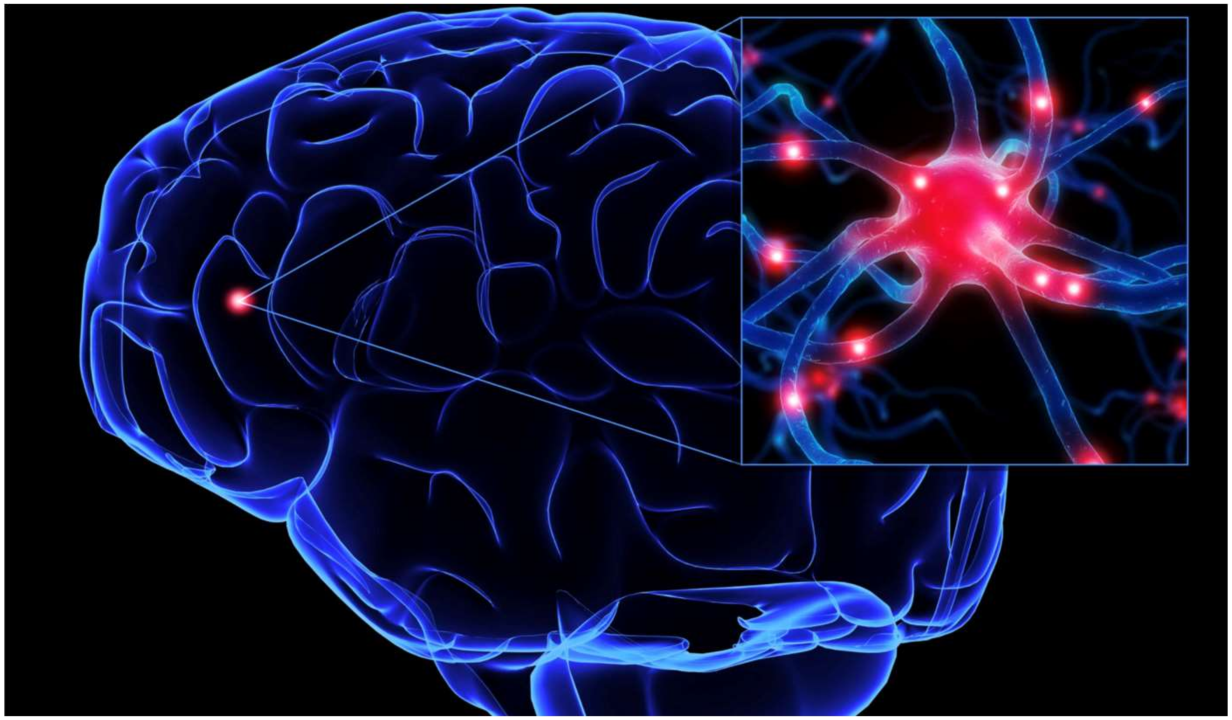
- Convolutional Neural Network
- Recurrent Neural Network
- ARIMA

Anomaly Detection

- Masked Autoencoder for Density Estimation
- Normal Hidden Markov Model

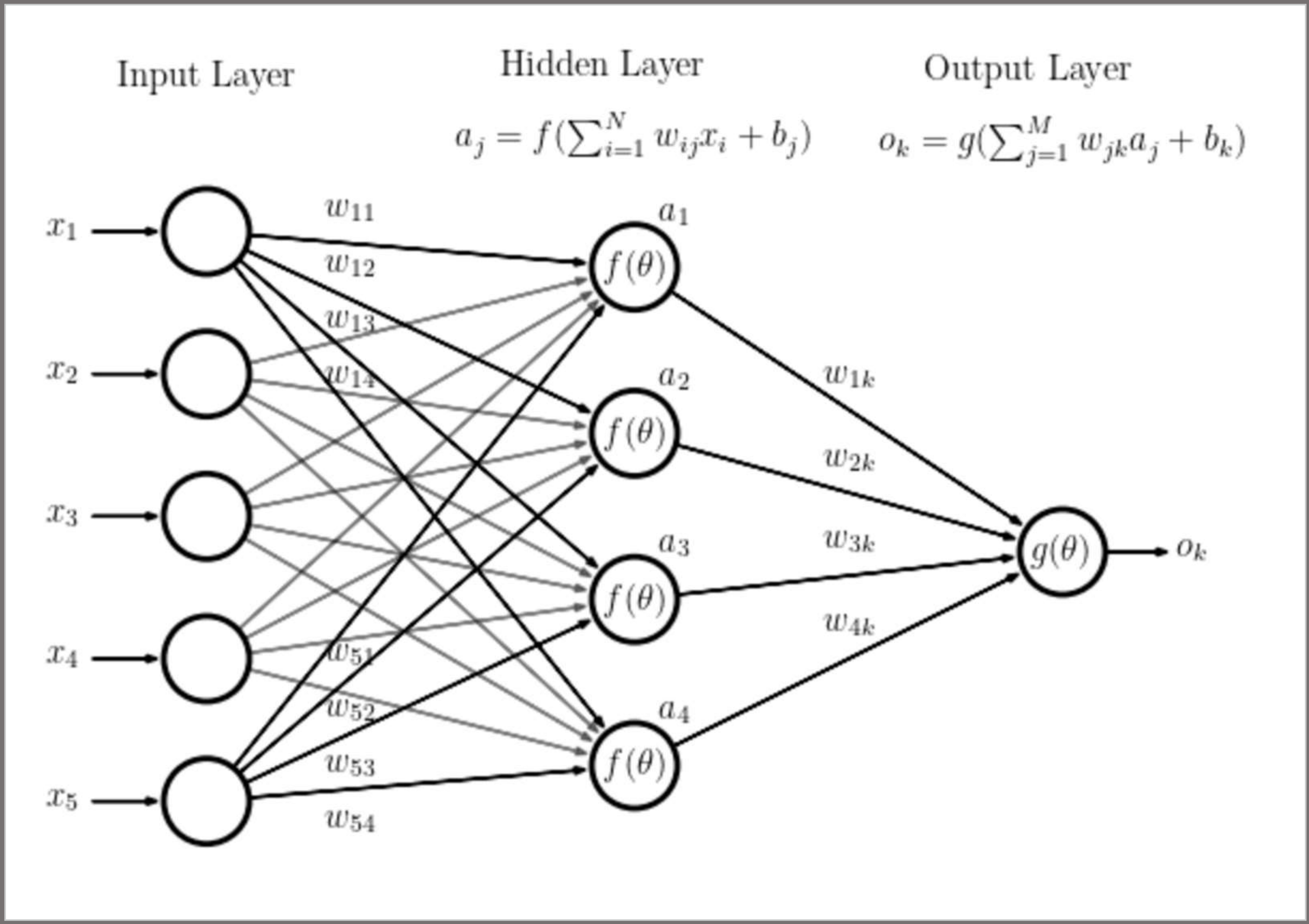
Research Questions

1. What architectural differences need to be made for a neural network to thrive with different data types and different tasks?
2. What are the strengths and weaknesses of neural networks in comparison to traditional statistical models?
3. In what situations would a neural network be preferred over simpler traditional models?

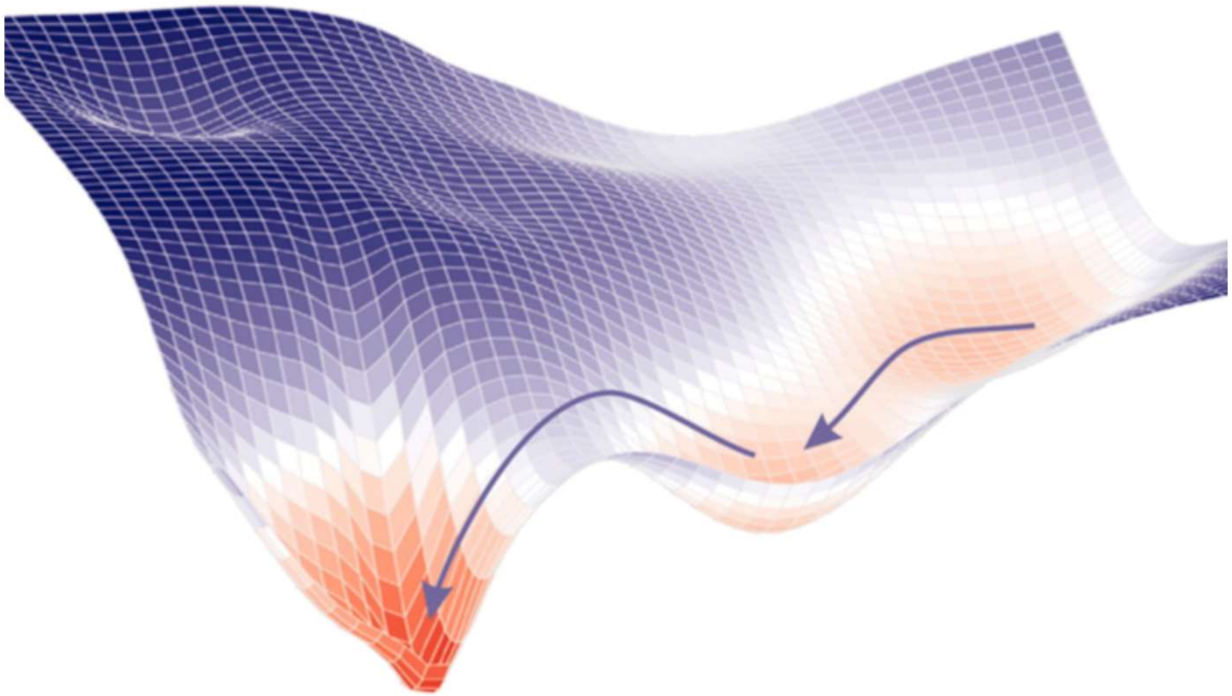


Neural Networks and the Human Brain

- Designed to model the way the human brain learns.
- Flexible models with interchangeable components: hidden units, layers, activation functions, loss functions.



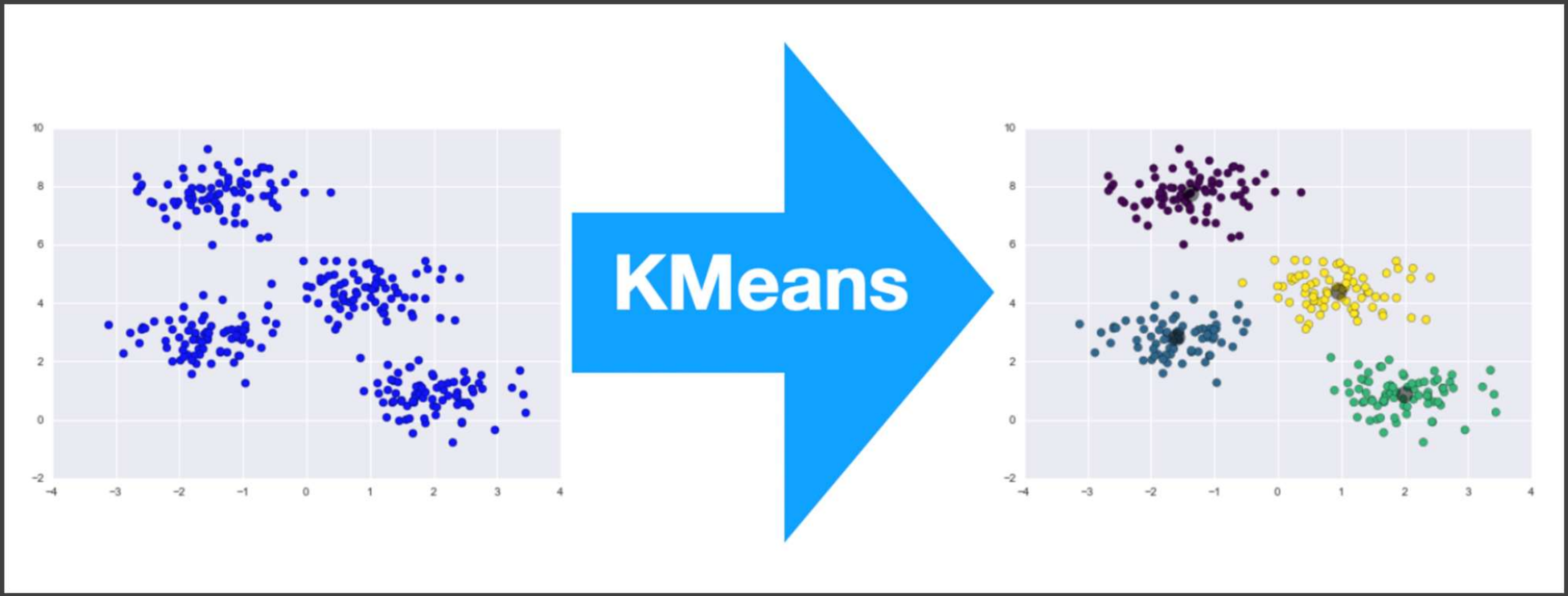
Neural Network Architecture Example



Fitting Neural Networks and Optimization Algorithms

- Gradient descent optimization minimizes the loss function by iteratively moving down the steepest slope.
- Adam optimization implements momentum, controlling the acceleration of descent, and exponential decay of the squared gradients.

Image Segmentation

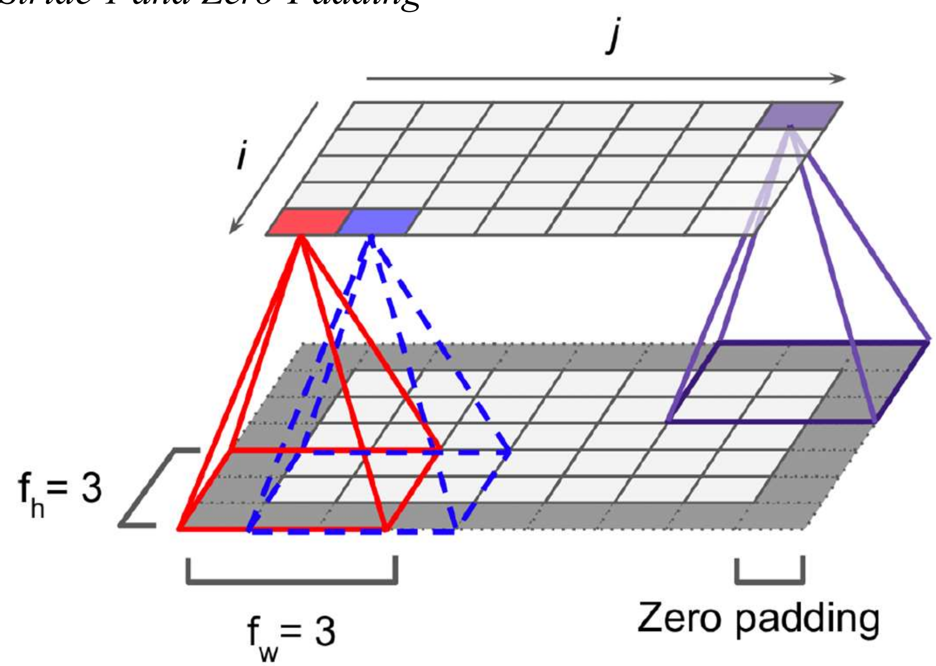


K-Means Clustering

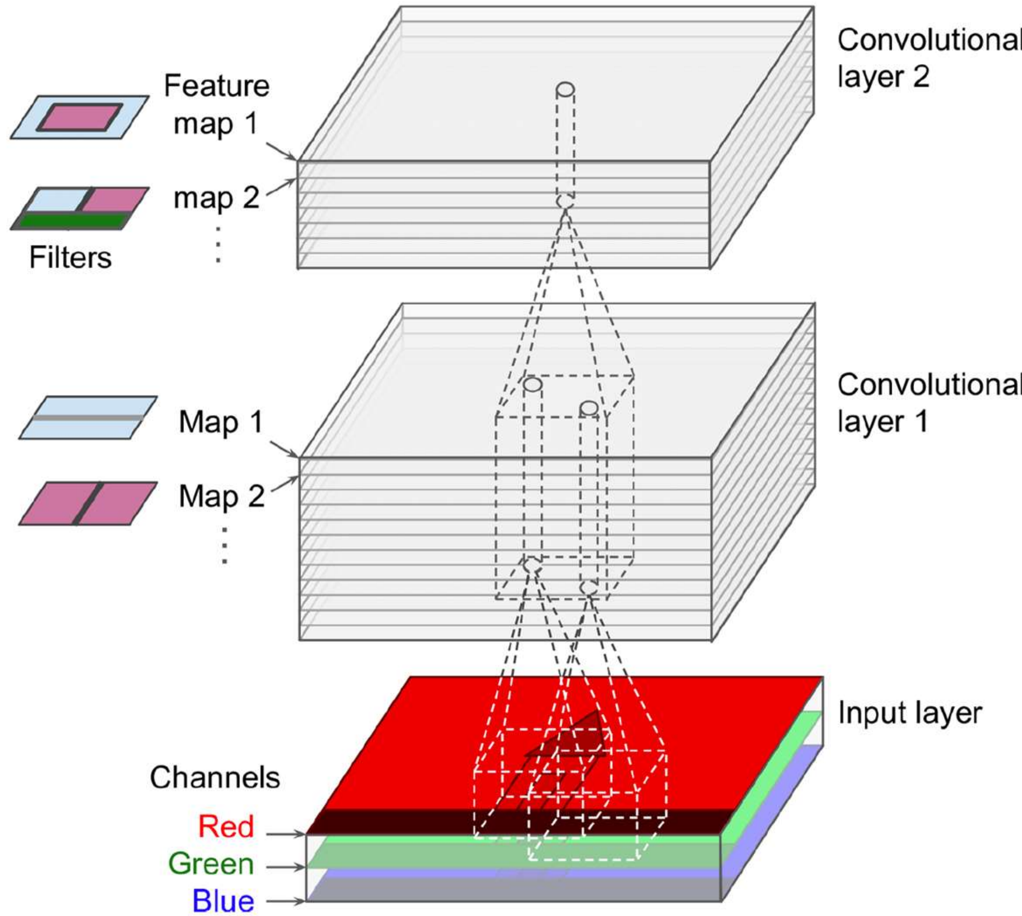
- K-Means clustering clusters the pixel values into K groups based on distance.
- For this application, $K = 2$.

Convolutional Layers

Convolutional Feature Map from 3x3 Filter with Stride 1 and Zero-Padding

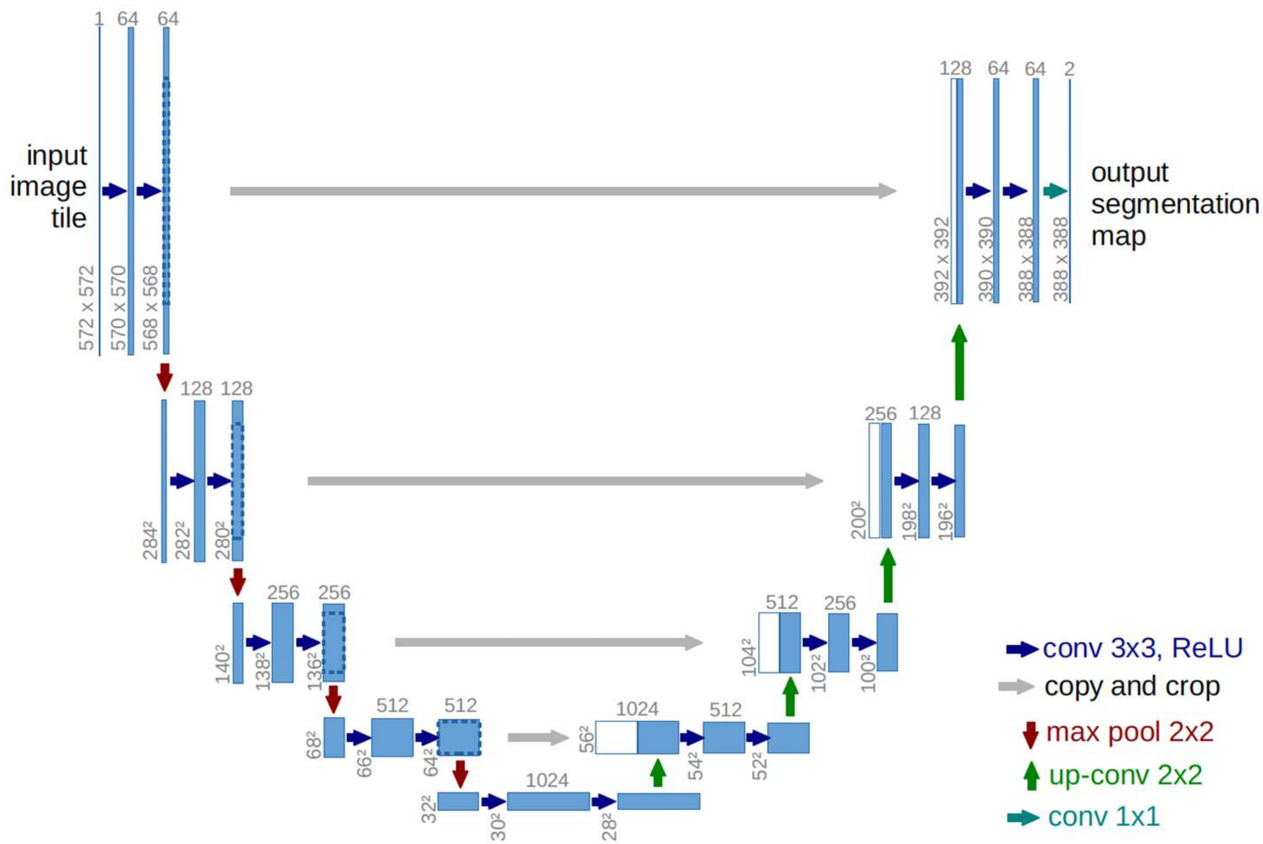


Stack of Two Convolutional Layers

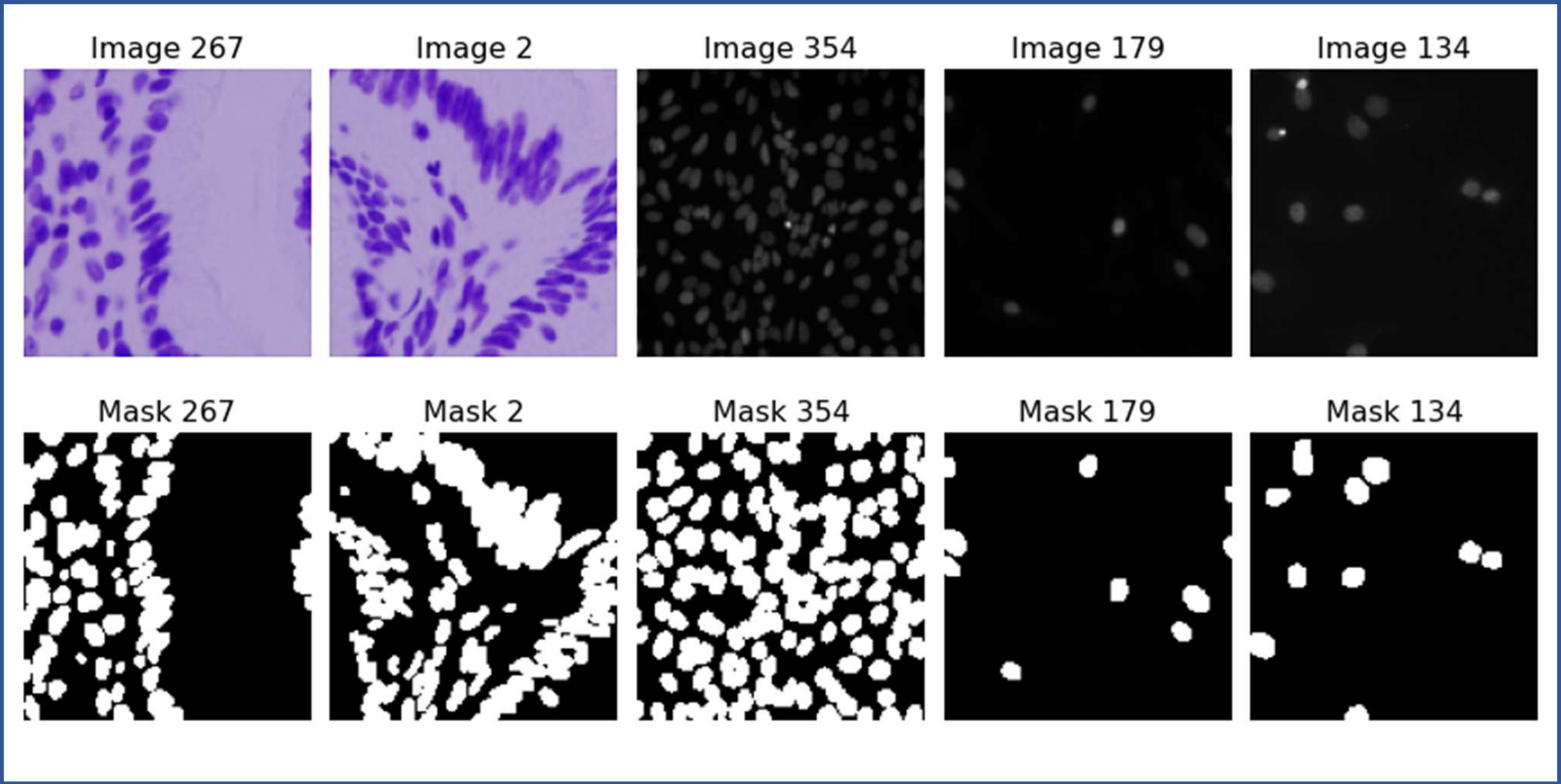


U-Net Fully Convolutional Network

- U-Net is a fully convolutional network with 31,031,745 trainable parameters designed for image segmentation.
- Outputs a segmentation mask.
- Architecture has skip connections from the contraction path to the expansion path.
- U-Net architecture example shown to the right.



Nuclei Dataset for Experimentation



Fitting the U-Net

- Google Collaboratory offers a free cloud-based GPU.
- Train data have 603 observations and 10% of train data are used for validation during training.
- The model is fit for 500 epochs, batch size 16, early stopping if no improvement in validation loss after 15 epochs.
- The model was fit for 98 epochs total that took 7-12 seconds each before training was stopped.

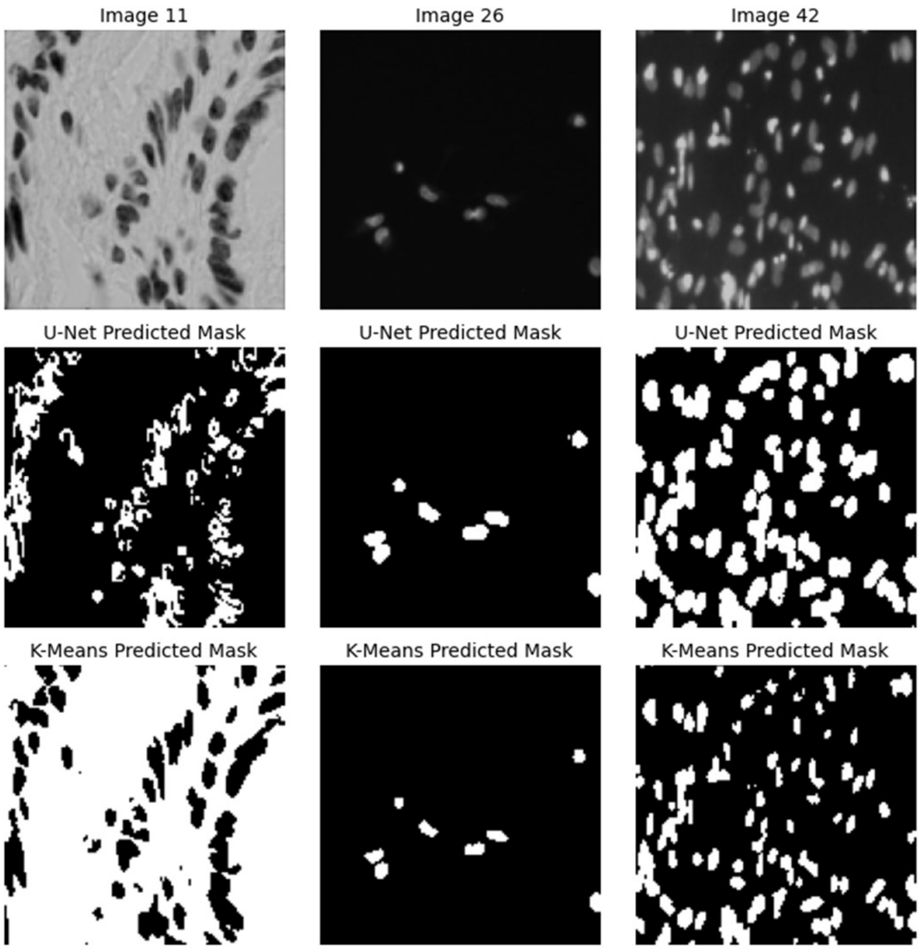
Model	Dice Coefficient	Soft Dice Loss
K-Means Train	0.1892	0.7332
K-Means Test	0.2008	0.7248
U-Net Train	0.9409	0.0591
U-Net Test	0.9312	0.0688

Evaluating Segmentation Models

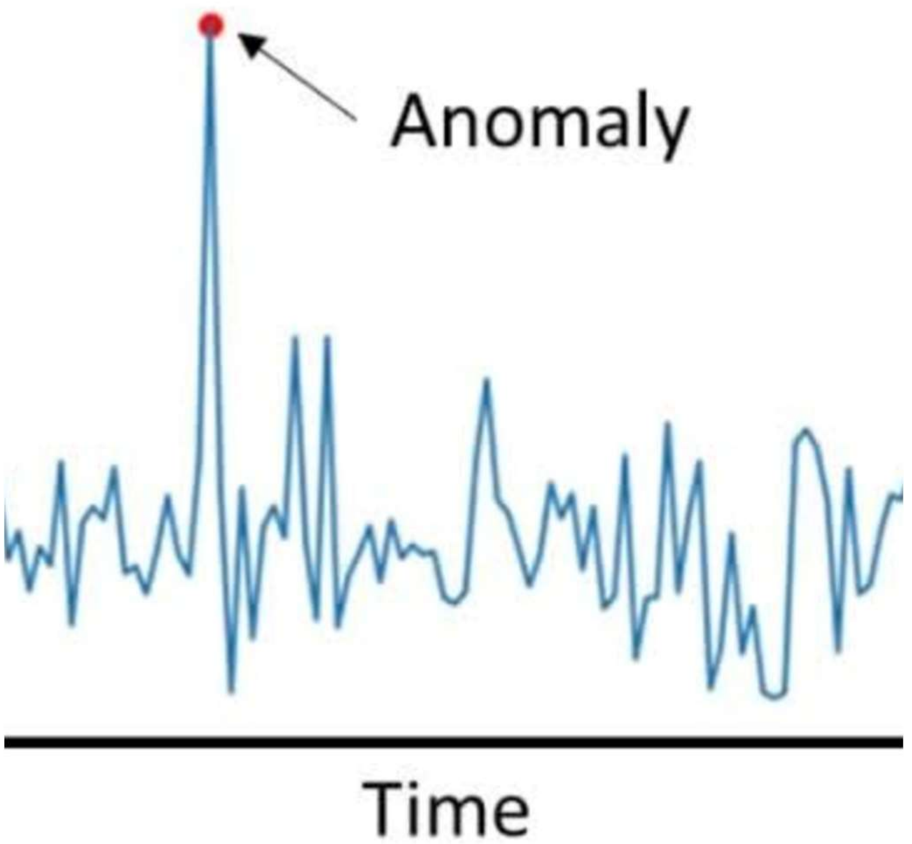
- U-Net performed very well on both train and test data based on performance metrics and is not overfit.
- K-Means clustering performed very poorly according to performance metrics, which will be discussed in the following slide.

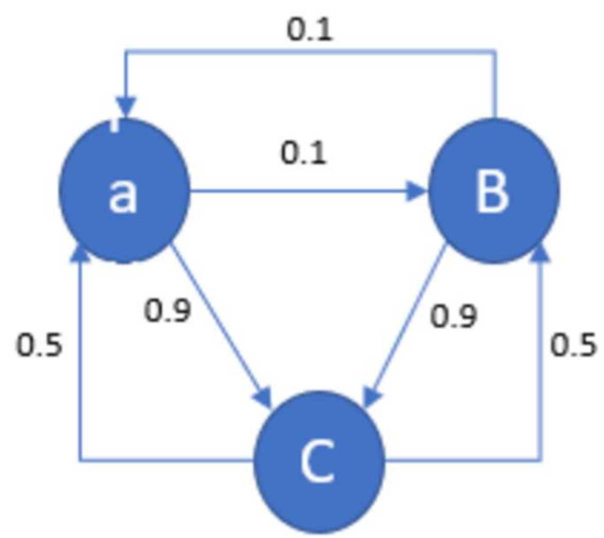
Predicted Masks for Test Observations

- U-Net predictions seem to be much more exact than K-Means.
- K-Means prediction for image 11 assigns pixels to the wrong class for segmentations.



Anomaly Detection for Sequential Data





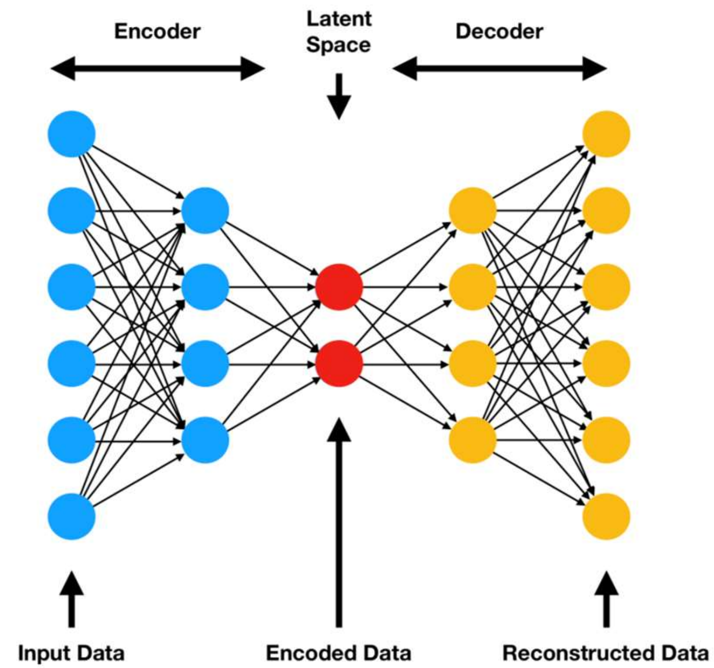
$$P = \begin{bmatrix} 0 & 0.1 & 0.9 \\ 0.1 & 0.0 & 0.9 \\ 0.5 & 0.5 & 0.0 \end{bmatrix}$$

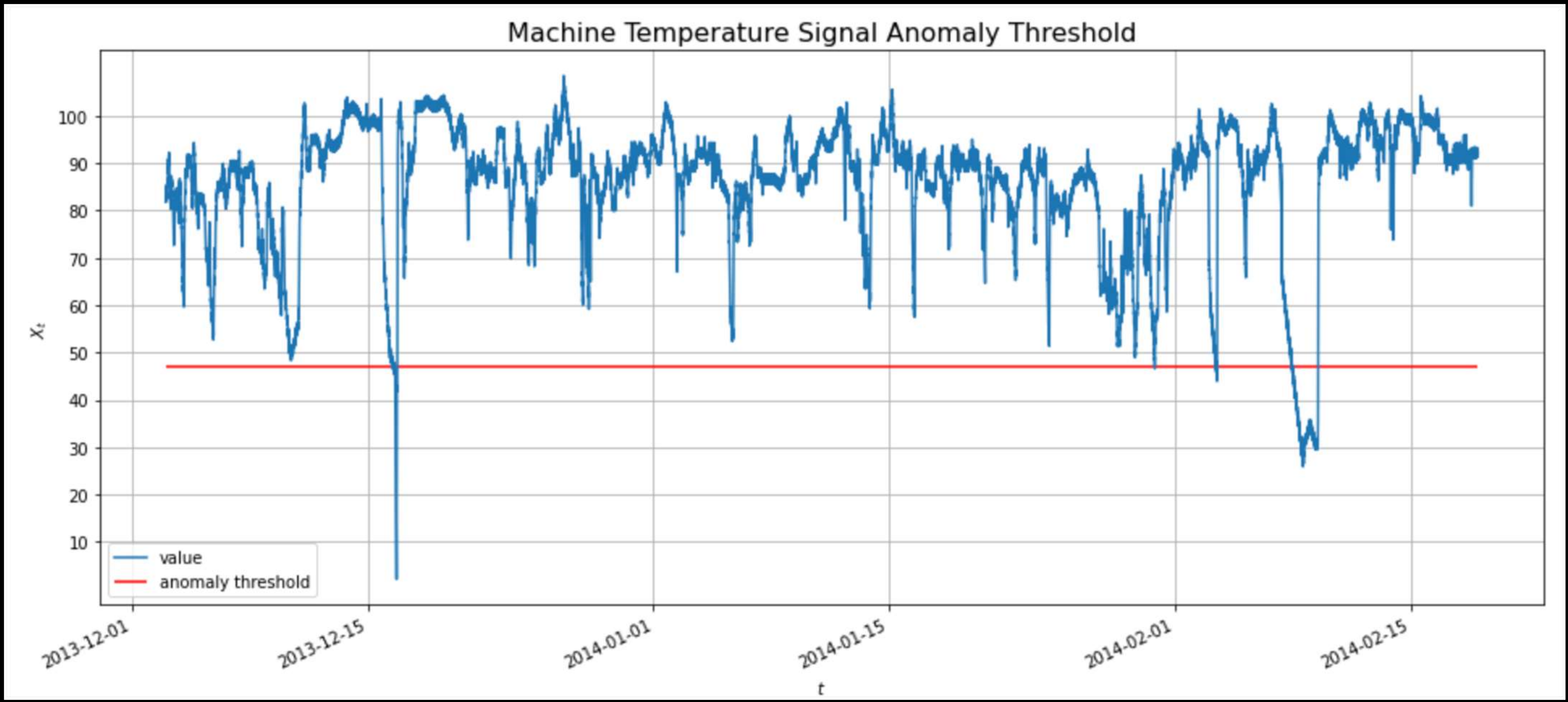
Hidden Markov Model

- Hidden Markov models assume a latent process considered to be a discrete Markov chain.
- The latent states of a hidden Markov model follow some probability distribution to generate observed values.

Masked Autoencoder for Density Estimation

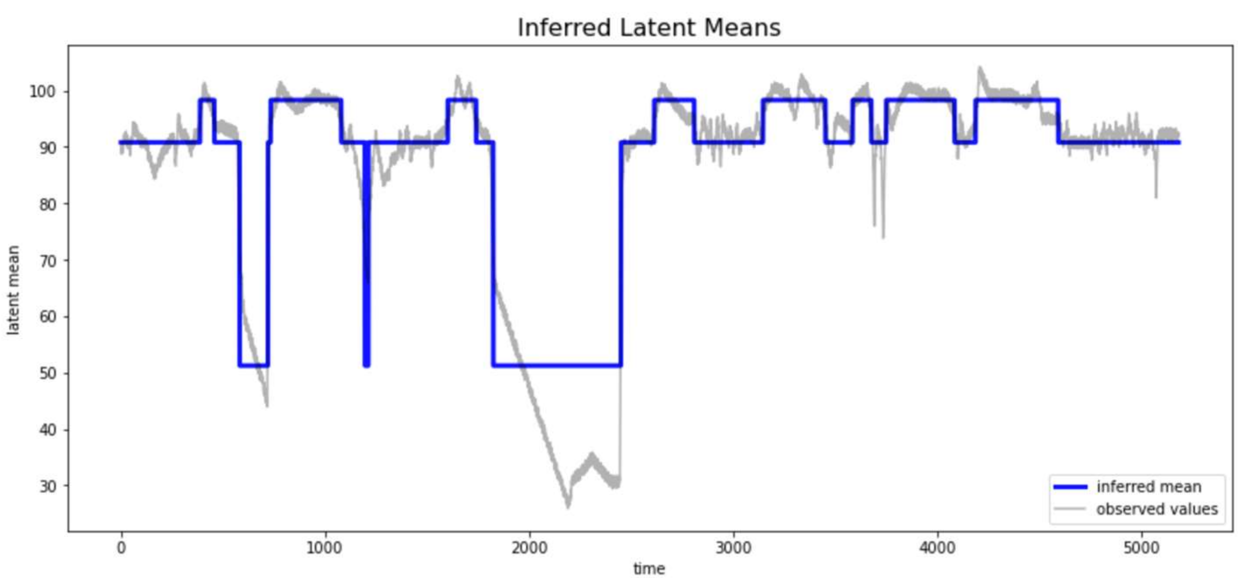
- Building block for the Masked Autoregressive Flow as a bijective transformation.
- Decomposes the joint density function into a product of one-dimensional conditionals.
- MADE uses binary masks to block any computational paths to future observations to enforce the autoregressive property.





Machine System Temperature Signal for Experimentation

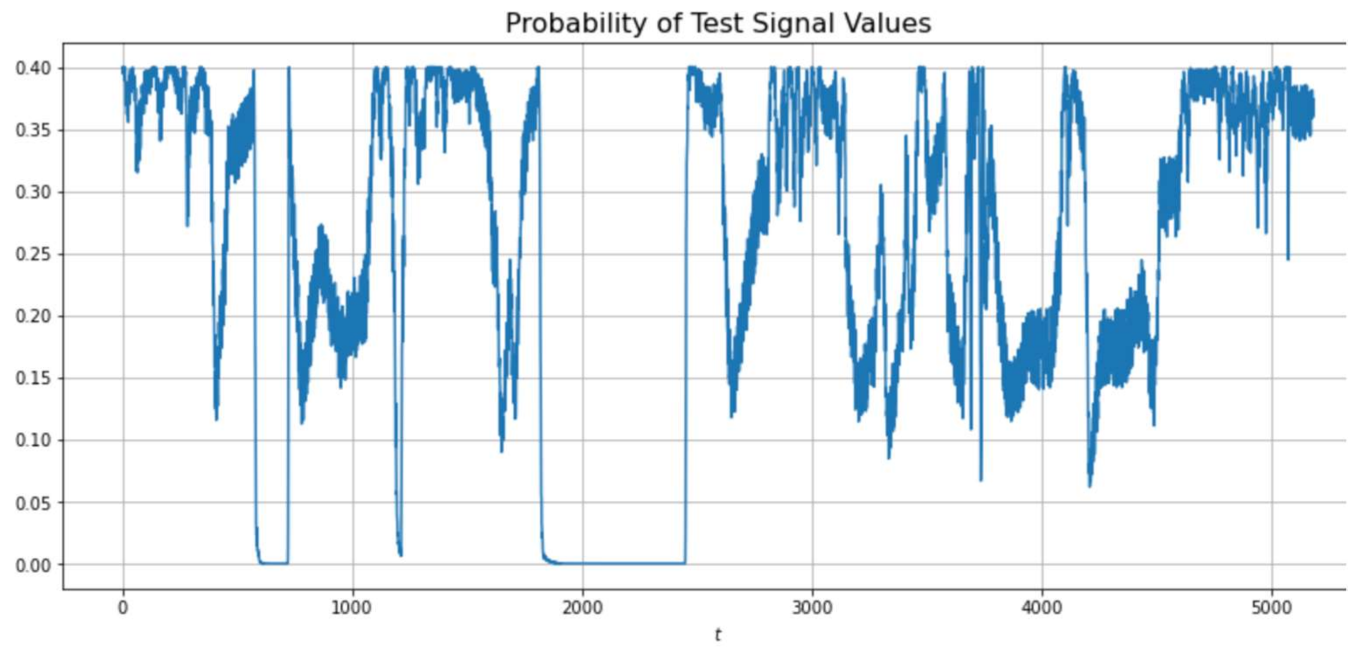
Normal Hidden Markov Model



- Observations follow a Normal distribution where the mean is conditional on the latent state.
- Variance is set constant for all states.
- HMM is fit with the Baum-Welch algorithm to compute the MAP fit to the observed data.
- HMM does not need to be initially fit on a train dataset.

Fitting MADE and Compute Test Probabilities

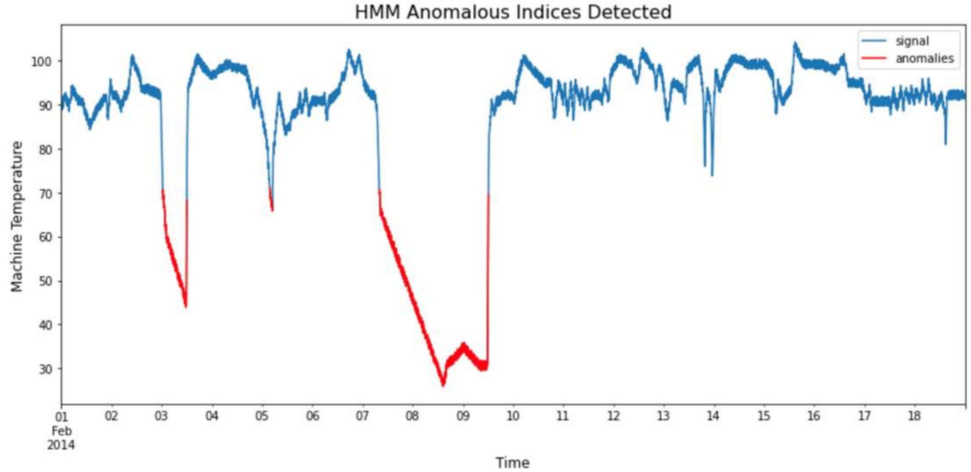
21

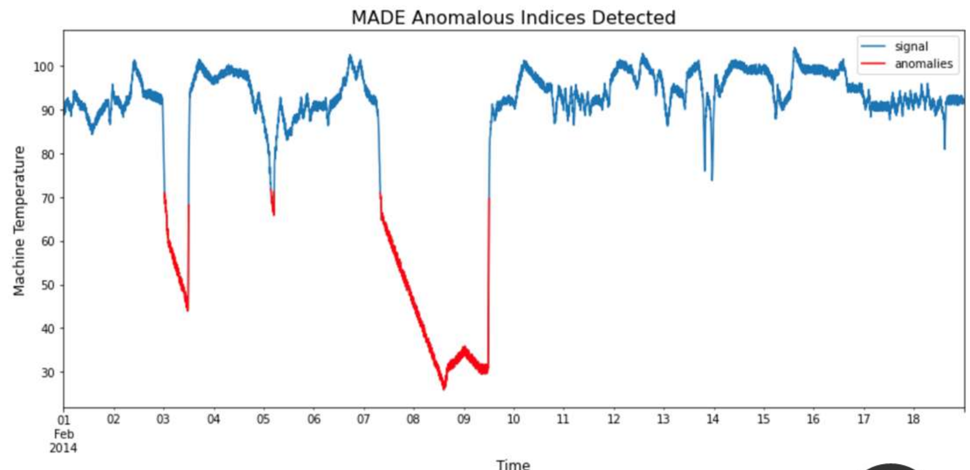


- MADE is a neural network and must be fit to a train dataset.
- Train dataset is determined as a portion of the signal with expected behavior.
- The model is fit for 100 epochs with early stopping to stop training after 5 epochs if there is no improvement in validation loss.
- The fitted model is a density function that can be used to compute the probabilities of the test signal values.

Detecting Anomalies

- In this application, anomalies have been defined as observations where the temperature drops below 47 degrees.
- **HMM**: anomalies are the observations that belong to the latent state with the minimum mean.
- **MADE**: anomalies are the observations that have a probability of occurring less than 0.05.





Model	Precision	Recall	F_1
HMM	0.7845	0.9644	0.7253
MADE	0.7827	0.9639	0.7224

Evaluating Anomaly Detections

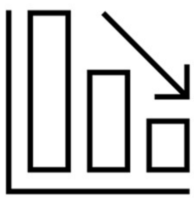
- The HMM and MADE produces almost indistinguishable results.
- The performance metrics and plotted detections look almost identical between the two models.
- Both models raised a false detection which is why the precision score is much lower than the recall scores.
- The false detections are much less costly than a missed detection because it is said that a missed detection results in a catastrophic system failure.

Research Question 1

What architectural differences need to be made for a neural network to thrive with different data types and different tasks?

- Neural networks are very flexible models with interchangeable components allowing it to work for many data types.
- The number and type of layers can be exchanged depending on the type of input data, allowing the layers to act as extensive statistical transformations that learn general features of the data.
- Activation functions for each layer also play a role in how the model performs since it transforms the entire output of a layer based on how the activations behave.
- The loss function must be selected so the learned parameters can be updated through the fitting procedures.

Research Question 2



What are the strengths and weaknesses of neural networks in comparison to traditional statistical models?

- Neural networks typically excel with large amounts of data.
- If data are sparse, the parameters can be initialized based on a pre-trained network that has been fitted to similar data.
- They are far more versatile since they can easily translate to structured data as well as signal data.
- Statistical models are generally designed to do one or a few tasks.
- Statistical models are usually far more interpretable than neural networks.

Research Question 3

In what situations would a neural network be preferred over a simpler traditional model?

- Neural networks frequently provide much stronger predictive power than statistical models and are designed to do a specific task at a very high level.
- Neural networks tend to be much more time consuming but also provide much more flexibility if the mathematical and statistical components of the model are selected correctly.
- Statistical models are generally far more interpretable and would be a better choice for inferences other than predictions.
- Statistical models are usually quicker to fit but require more preprocessing steps, they will perform well when assumptions are met.

Concluding Thoughts

- This study provides a small preview of the capabilities of neural networks.
- Neural networks are a flexible class of models that can perform well with many data types.
- Selecting the right model entails much more than whether a model can perform a task or not.
- Neural networks provide a model option for most tasks.
- Neural networks will almost surely continue to grow in popularity as computational resources increase and research advances.

To the
Southern California Chapter of the American Statistical Association and all attendees,

Thank you for listening!!
I hope you enjoyed.

Joseph Isamu Gadbois

Email: josephgadbois17@gmail.com