Subsurface Data Analytics and Machine Learning

Machine Learning Examples

Lecture outline . . .

- Advanced Workflows
- Support Vector machine
- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- Long Short-Term Memory Networks (LSTM)

Introduction

Data Analytics

Inferential Methods

Predictive Methods

Advanced Methods

Conclusions

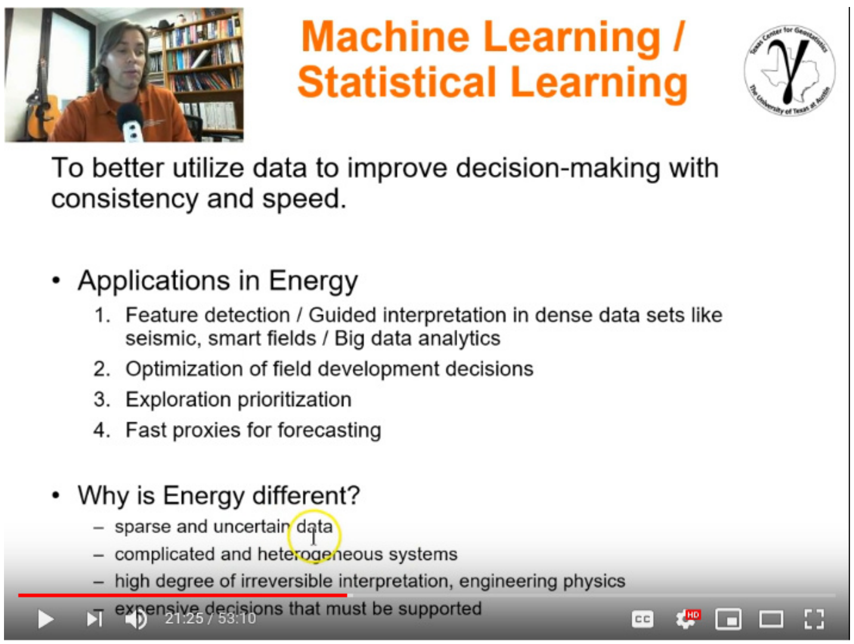
Instructor: Michael Pyrcz, the University of Texas at Austin

Subsurface Data Analytics and Machine Learning

Analytics and Machine Learning

Other Resources:

- **Statistical Learning, Dimensional Reduction and Decision Tree**



Instructor: Michael Pyrcz, the University of Texas at Austin

Goals of This Lecture

- Build awareness.

- Show opportunities in subsurface modeling with machine learning.

Subsurface Data Analytics and Machine Learning

Machine Learning Examples

Lecture outline . . .

- Advanced Workflows

Introduction

Data Analytics

Inferential Methods

Predictive Methods

Advanced Methods

Conclusions

Instructor: Michael Pyrcz, the University of Texas at Austin

Machine Learning for Subsurface Modeling

- A set of powerful machine learning methods to support subsurface modeling with initial workflows.
- We are actively working in this area.
- Concepts are represented in a simplified manner.
- We have workflows available.

Methods Covered:

- Support Vector Machines (SVM)
- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- Deep Convolutional Generative Adversarial Networks (DCGANs)

Subsurface Data Analytics and Machine Learning

Machine Learning Examples

Lecture outline . . .

- Support Vector machine

Introduction

Data Analytics

Inferential Methods

Predictive Methods

Advanced Methods

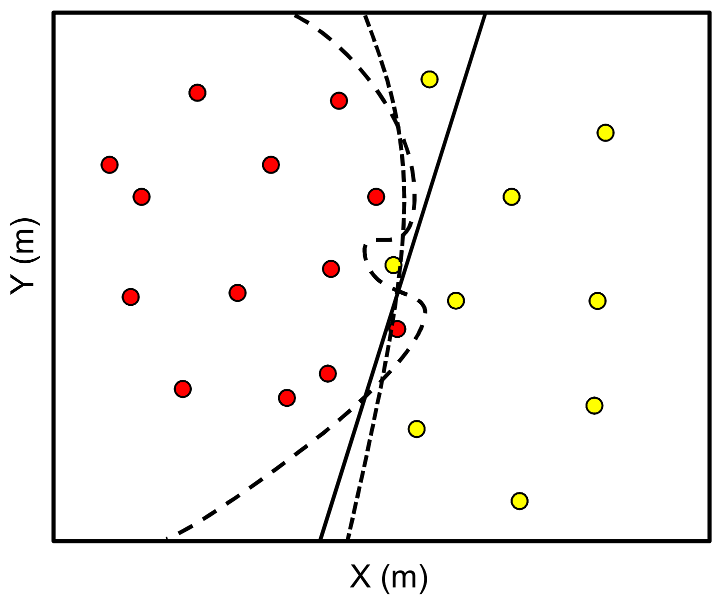
Conclusions

Instructor: Michael Pyrcz, the University of Texas at Austin

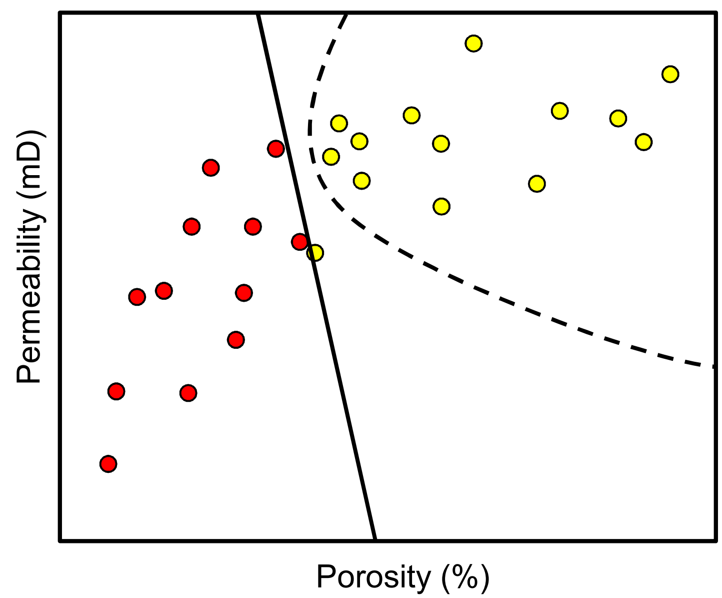
Support Vector Machines

- A powerful supervised training, machine learning method for segmentation
- For example, forming a rule to segment a multivariate dataset into multiple categories with a decision rule.
- E.g. Geoscience Kanevski et al., 2000, Bio Goovaerts et al., 2018

Spatial Boundaries



Multivariate Boundaries



Support Vector Machines

1. Form a boundary with the largest possible **margin** between the different cases.
2. Data within the margin update the model, they are called **support vectors**.
3. Project into problem into a higher dimensional space to solve linearly, with a variety of **kernels**.
4. The **C** parameter controls to penalty of misclassification, high C will result in a more complicated model (lower model bias, higher model variance.

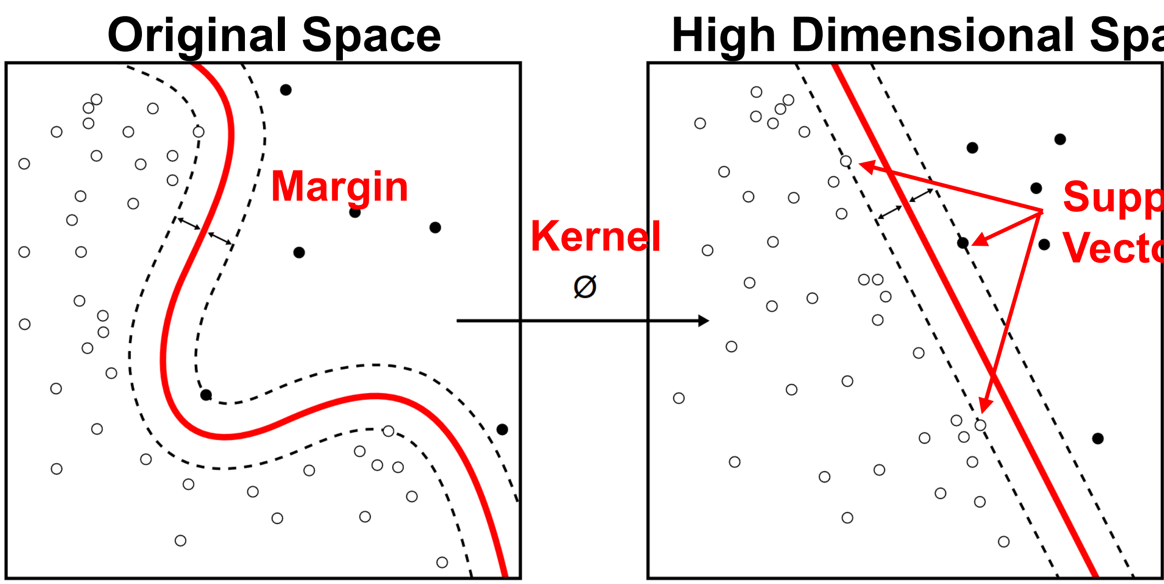


Image by Zirguezi, available at https://en.wikipedia.org/wiki/Support-vector_machine#/media/File:Kernel_Machine.svg

Support Vector Machines

Demonstration workflow with support vector machines to form a decision rule for segmentation.

Subsurface Data Analytics

Support Vector Machine for Multivariate Segmentation of Facies in Python

Wendi Liu, Michael Pyrcz, University of Texas at Austin

Workflow Goals

Learn the basics of support vector machine in python to segment facies given petrophysical properties. This includes:

- Loading and visualizing sample data
- Trying out support vector machine with different kernels (linear, polynomial, radial basis function)
- Tuning the SVM model parameters and results evaluation

Objective

I want to provide hands-on experience with building subsurface modeling workflows. Python provides an excellent vehicle to accomplish this.

The objective is to remove the hurdles of subsurface modeling workflow construction by providing building blocks and sufficient examples. This is not a coding class per se, but we need the ability to 'script' workflows working with numerical methods.

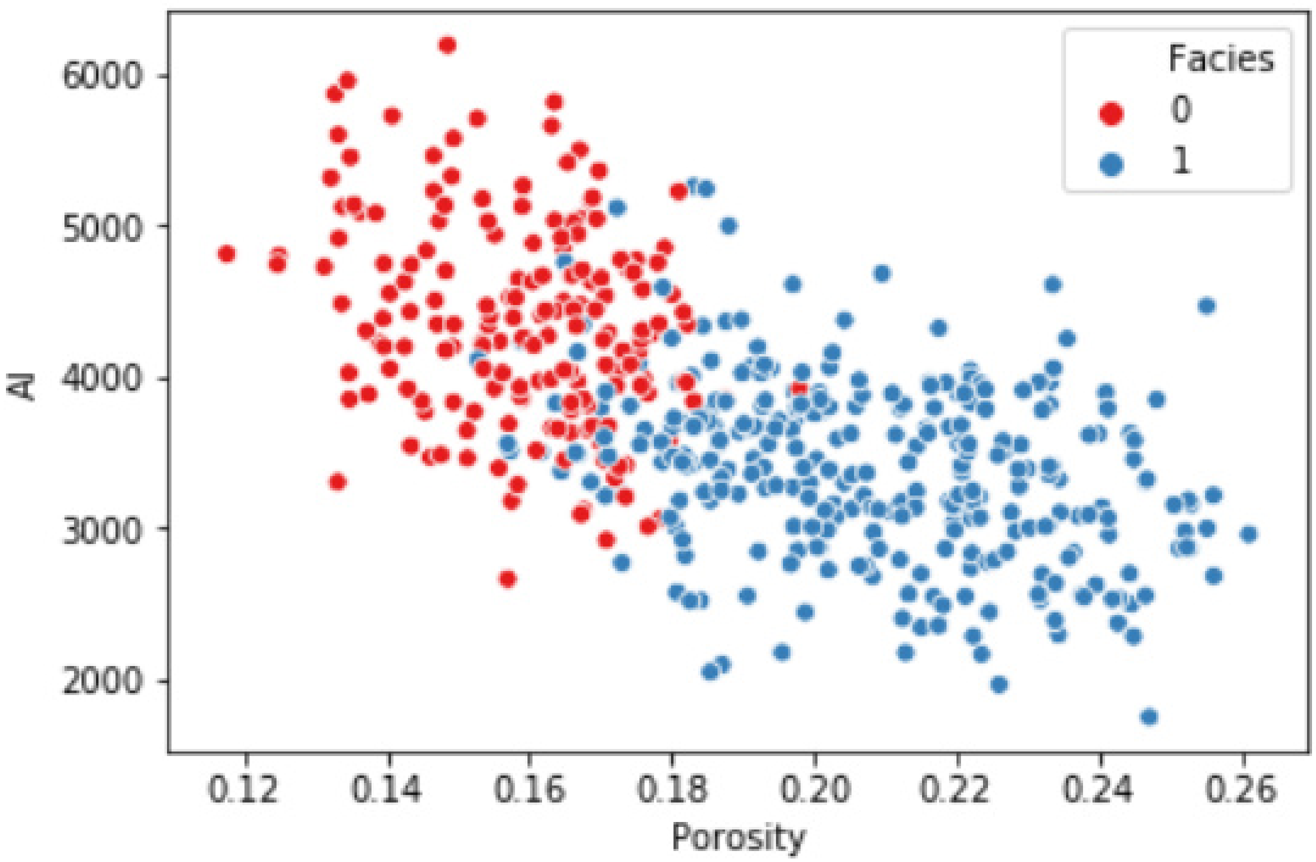
Load the required libraries and functions

The following code **imports** the required libraries. After we excute this code we can use 'os', 'np', 'pd' to access functionality in each of these libraries.

Image by Zirguezi, available at https://en.wikipedia.org/wiki/Support-vector_machine#/media/File:Kernel_Machine.svg

Support Vector Machines

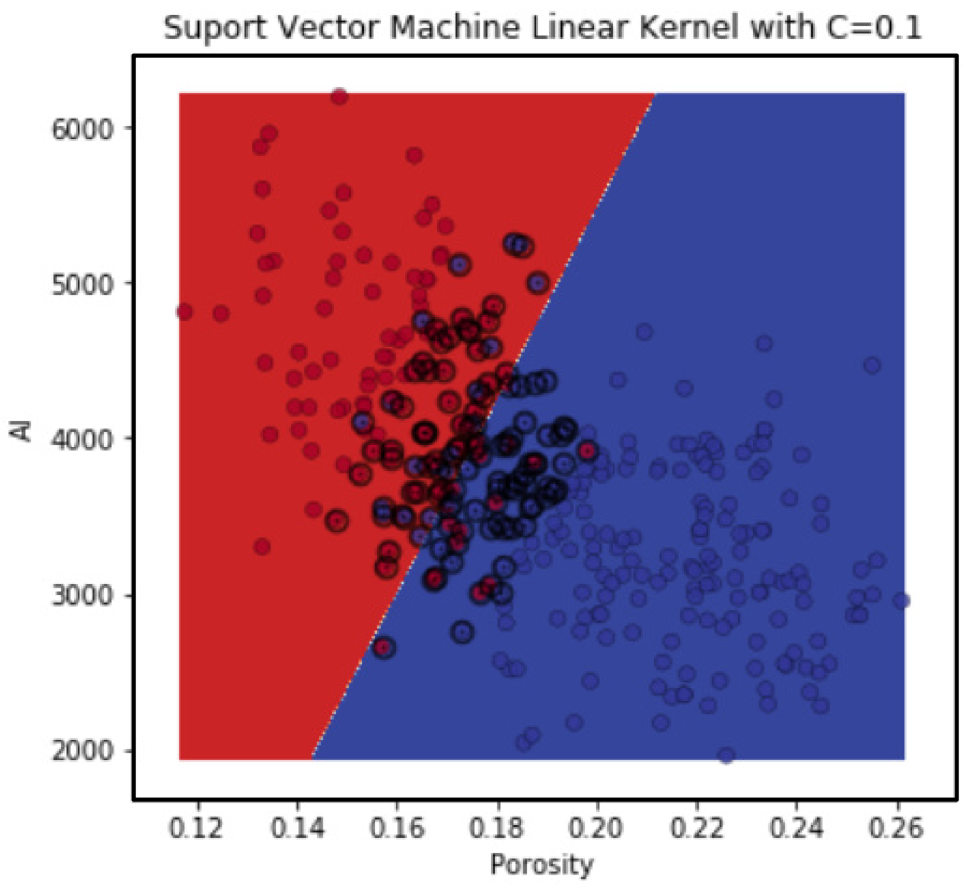
Demonstration workflow with support vector machines to form a decision rule for segmentation. Here's the training data.



Workflow developed by Wendi Liu, PhD student at The University of Texas at Austin.

Support Vector Machines

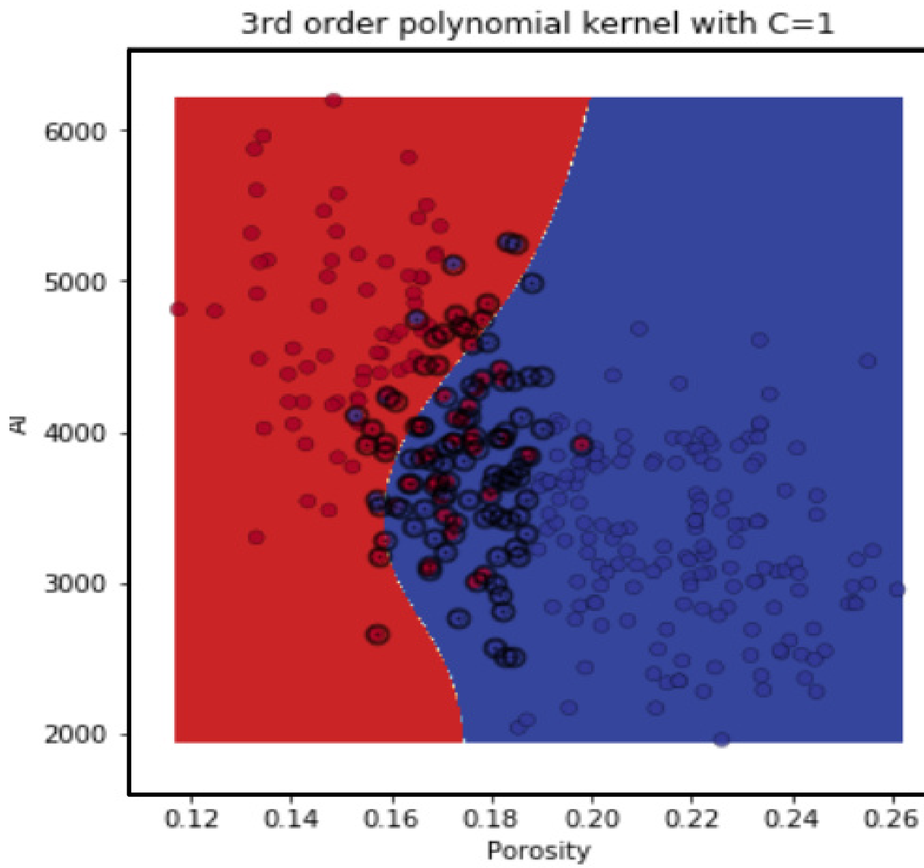
Linear kernel – decision boundary and support vectors highlighted.



Workflow developed by Wendi Liu, PhD student at The University of Texas at Austin.

Support Vector Machines

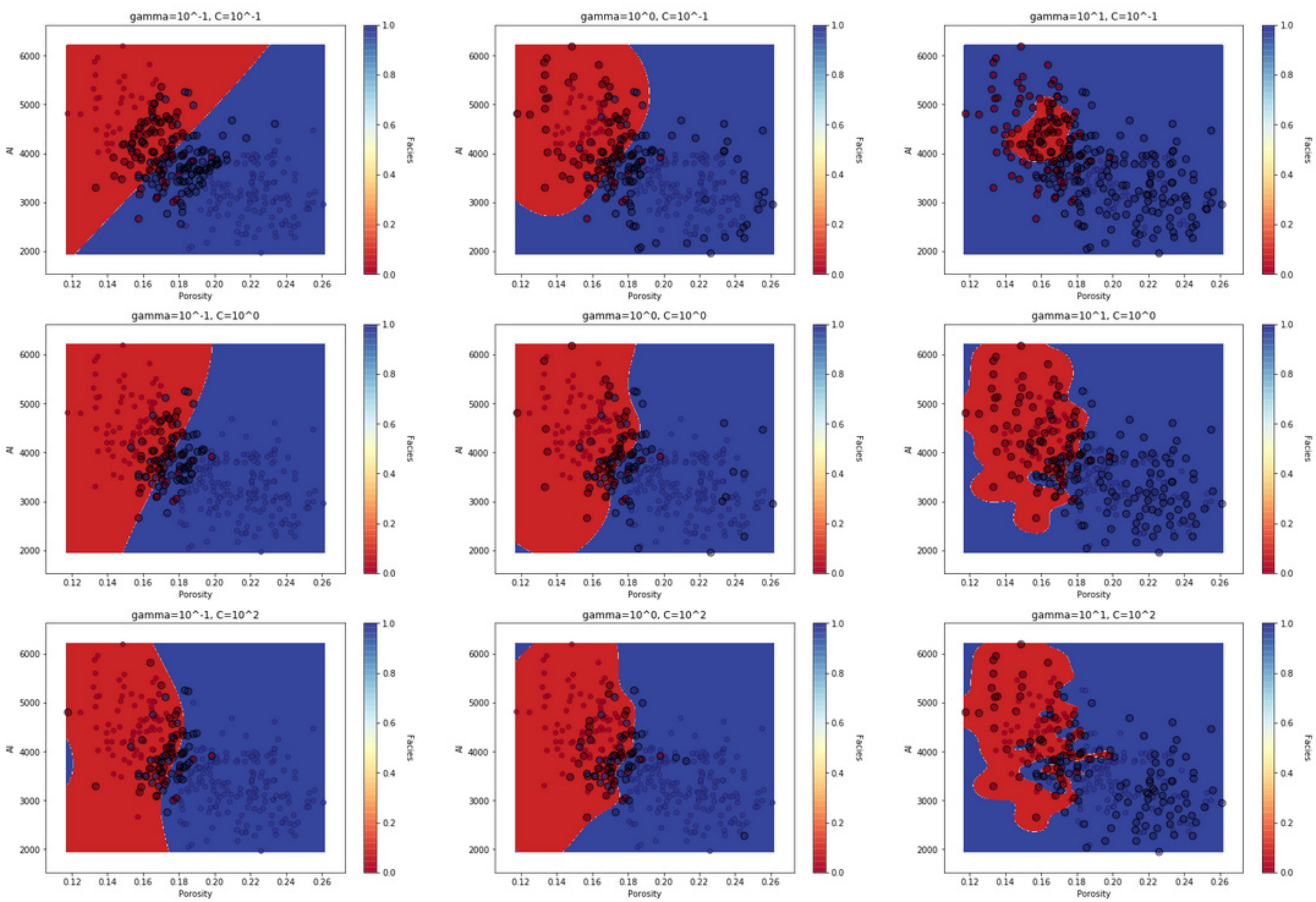
3rd Order Polynomial Kernel – decision boundary and support vectors highlighted.



Workflow developed by Wendi Liu, PhD student at The University of Texas at Austin.

Support Vector Machines

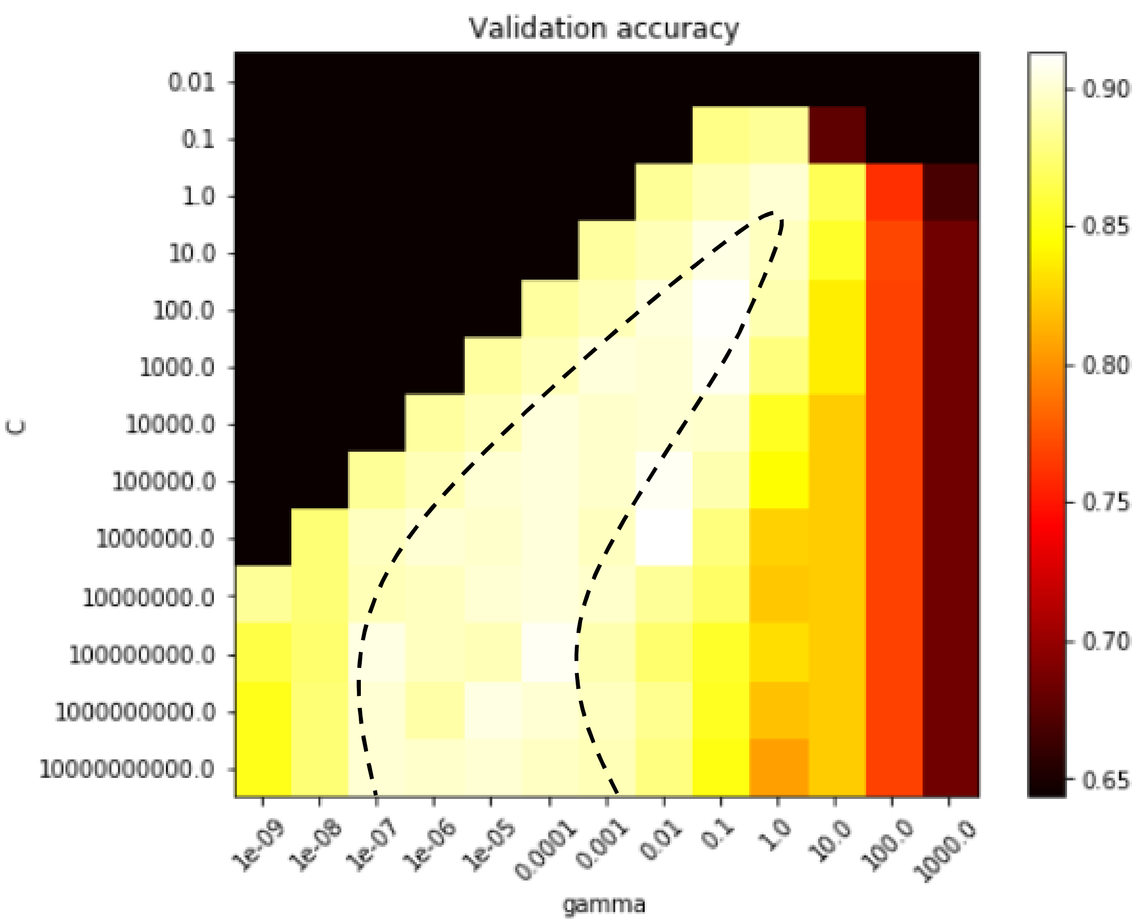
Radial Kernel – control of C and curvature parameter, γ.



Workflow developed by Wendi Liu, PhD student at The University of Texas at Austin.

Support Vector Machines

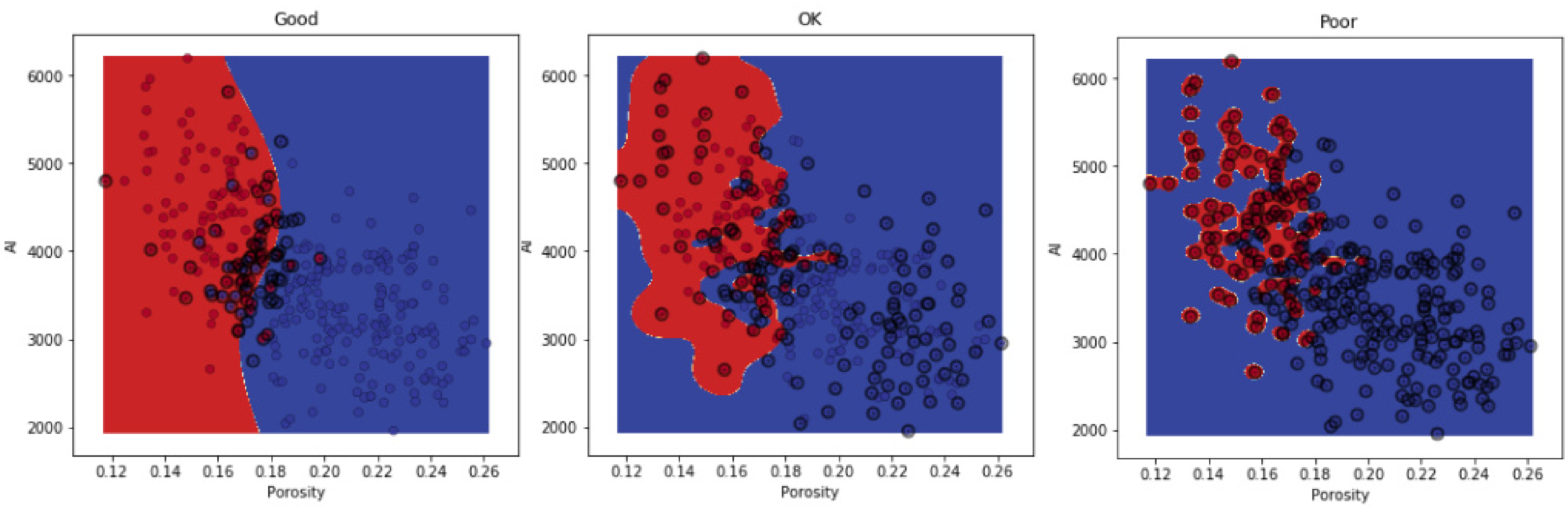
Performance with testing for a wide variety of parameters.



Workflow developed by Wendi Liu, PhD student at The University of Texas at Austin.

Support Vector Machines

Examples of models that perform well, average and poorly in testing.



- A clear case of overfit.

Workflow developed by Wendi Liu, PhD student at The University of Texas at Austin.

Subsurface Data Analytics and Machine Learning

Machine Learning Examples

Lecture outline . . .

- **Advanced Workflows**
- **Artificial Neural Networks (ANN)**

Introduction

Data Analytics

Inferential Methods

Predictive Methods

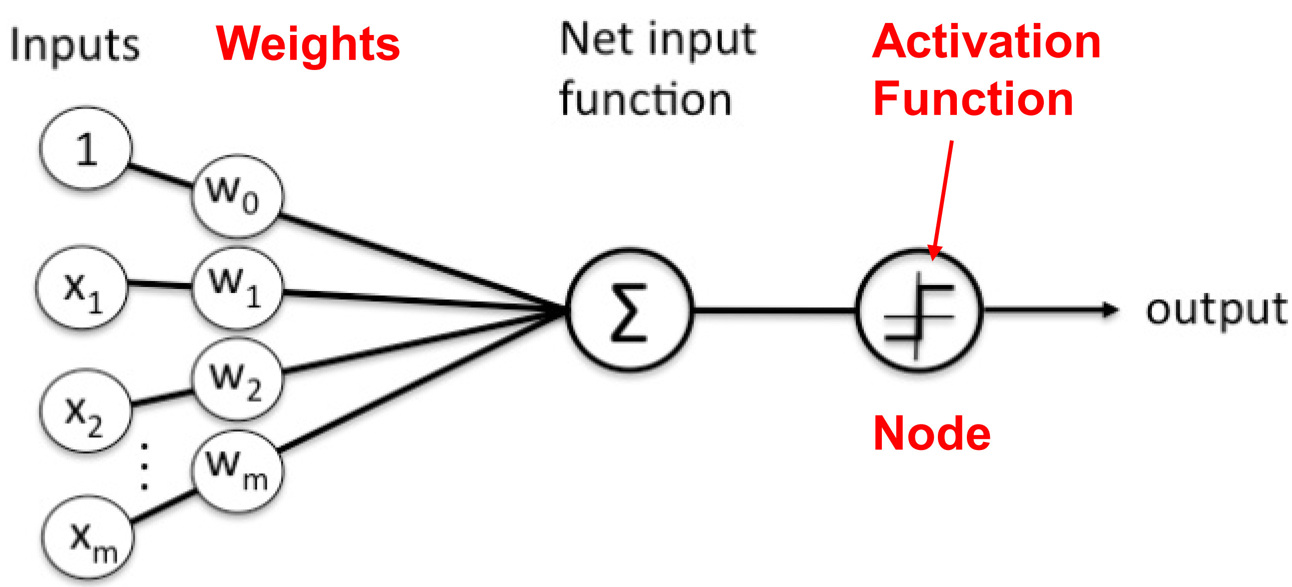
Advanced Methods

Conclusions

Instructor: Michael Pyrcz, the University of Texas at Austin

Neural Nets

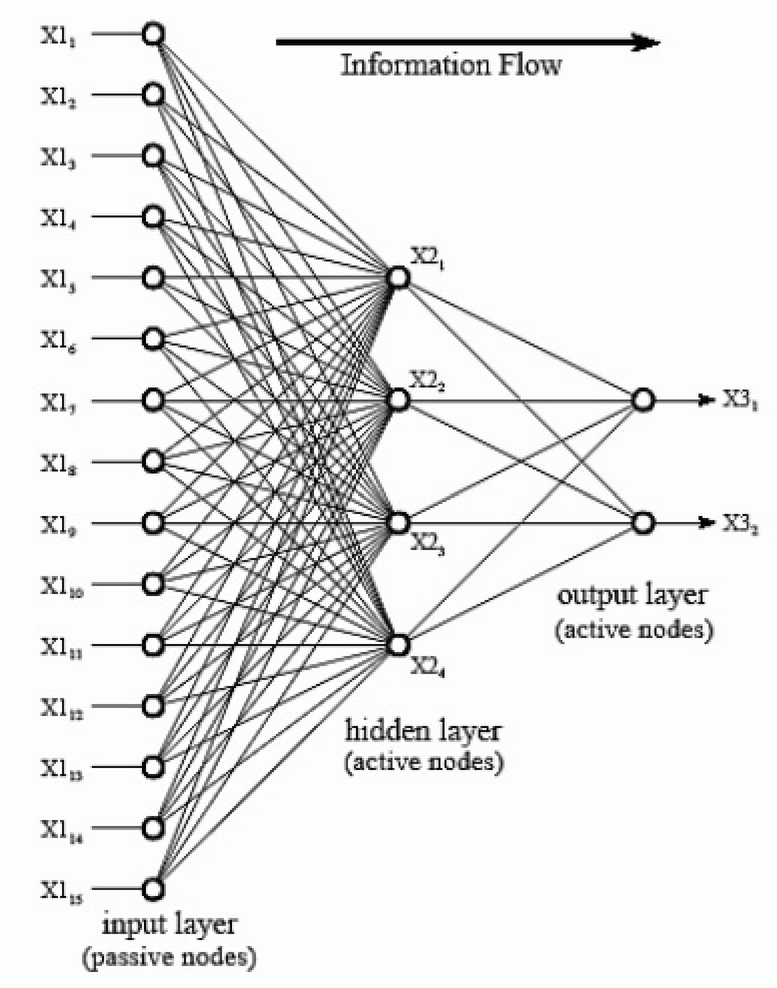
- Mimic the human brain for pattern recognition
- A set of interconnected **nodes**, organized into layers
- Nodes fire when their input exceeds a threshold, **activation function**
- Each data-to-node or node-to-node path has **weights**
- Weights and activation functions are **trained** to improve accuracy



Single Node Neural Net Image from https://skymind.ai/wiki/neural-network

Neural Nets

- As the number of nodes increases along with the number of node layers and interconnections, more complicated predictions are possible.
- Complicated activation functions and signals are also possible.
- These are very parameter rich, complicated models with low model bias but high model variance.
- Require more training data.
- Deep learning simply means more than one hidden layer!



Fully connected neural net diagram from https://www.dspguide.com/ch26/2.htm

Neural Nets

- Demonstrate neural nets with MPS training images and simulated models
- Train a neural net to determine the TI that resulted in the current simulated model

Workflow developed by Honggeun Jo, PhD student at The University of Texas at Austin.

Neural Nets

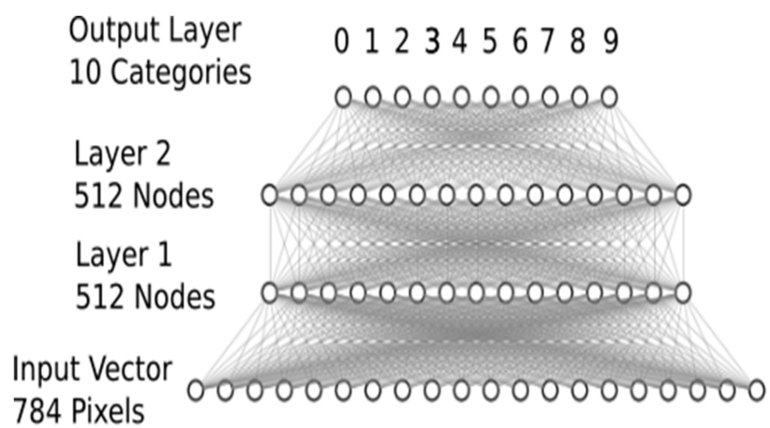
- Generate TIs with 9 different setting
  (4 channels, 4 lobate shapes, and 1 circular sand body)
- Realize 1,000 models each from different Tis
- Label their realization with the related Ti
- Construct Neural Networks (ANN)
  - Input: reservoir (facies) model
  - output: probability to be realized from each TI (different geo. set)

Workflow developed by Honggeun Jo, PhD student at The University of Texas at Austin.

Neural Nets

- Modified National Institute of Standards and Technology database
- 60,000 training and 10,000 testing images of digits with labels

- Optical digit recognition is an artificial neural net



Number from 0 - 9

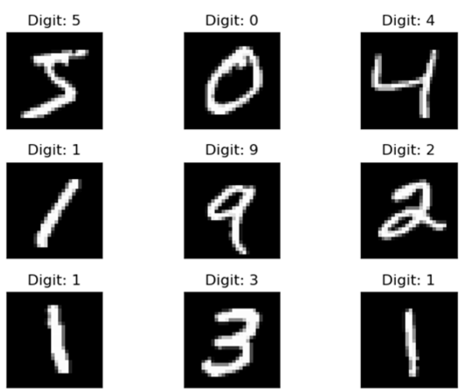
Value in each pixel.

MNIST image from https://upload.wikimedia.org/wikipedia/commons/2/27/MnistExamples.png

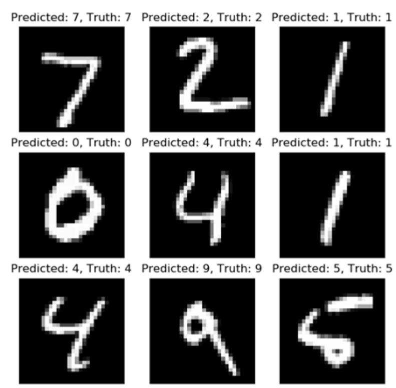
Neural Nets

- Modified National Institute of Standards and Technology database
- 60,000 training and 10,000 testing images of digits with labels

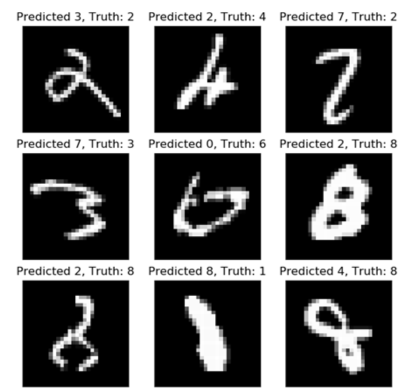
Training



Correct Classification

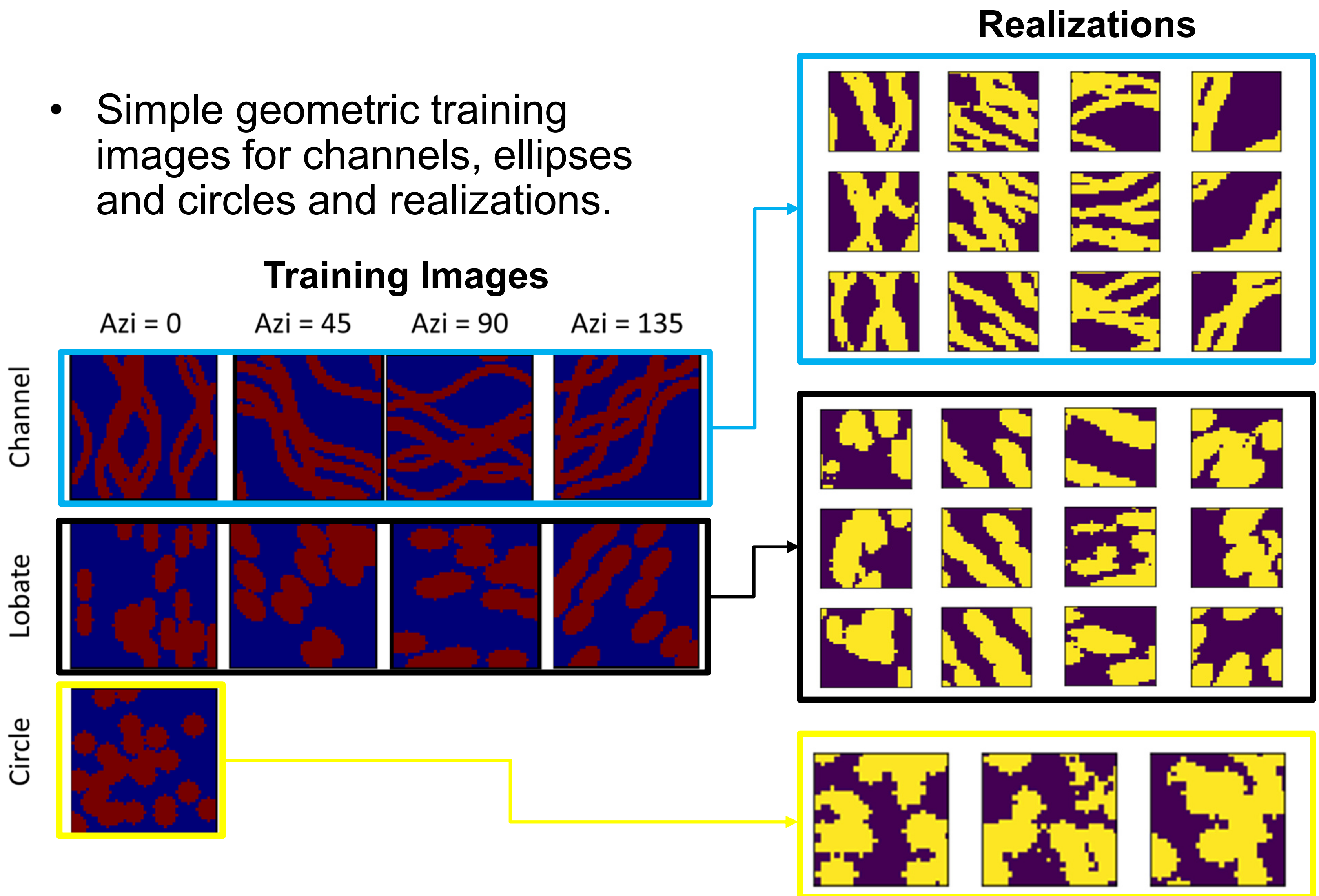


Incorrect Classification



Neural Nets

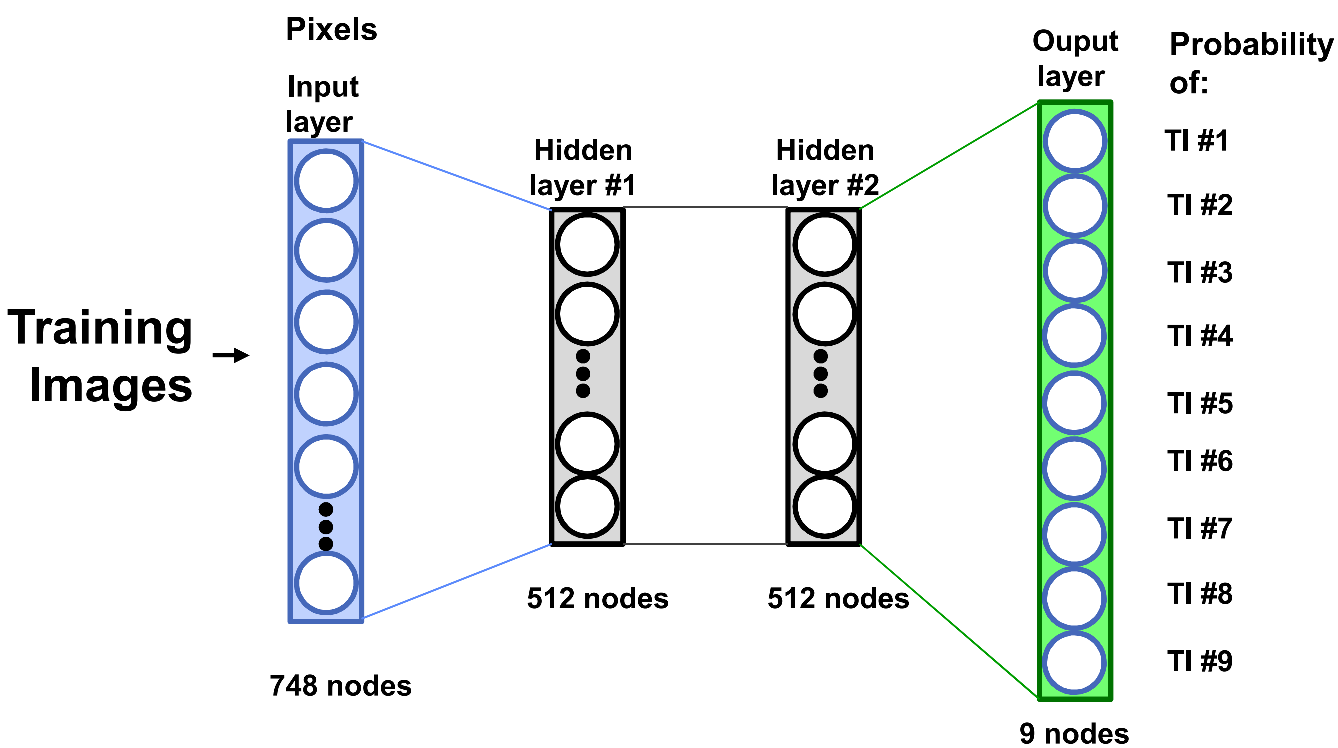
- Simple geometric training images for channels, ellipses and circles and realizations.



Workflow developed by Honggeun Jo, PhD student at The University of Texas at Austin.

Neural Nets

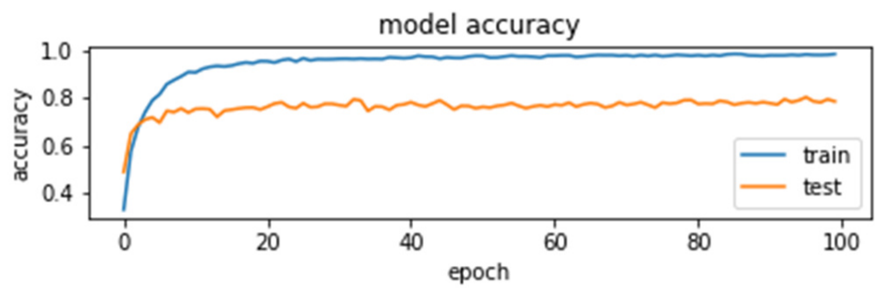
- Structure of the artificial neural network.
- Note, the image is ‘flattened’ to a list of pixel values.



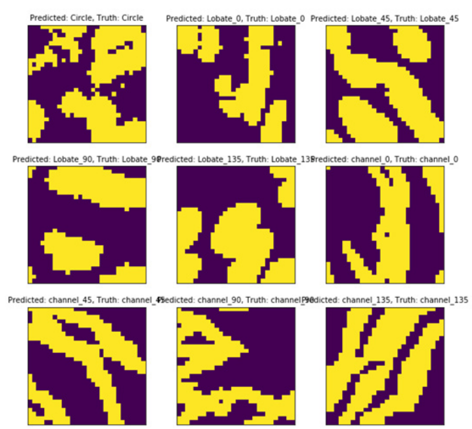
Workflow developed by Honggeun Jo, PhD student at The University of Texas at Austin.

Neural Nets

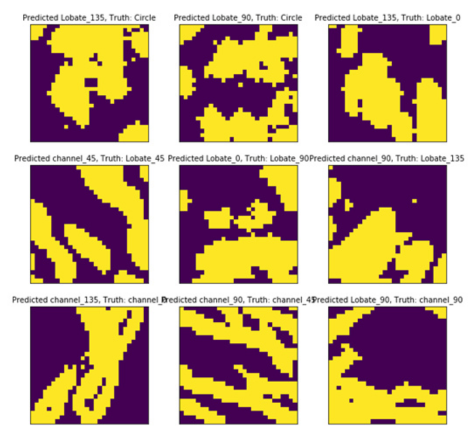
- The training and testing accuracy vs. number of training cycles, **Epoch**
- Levels off at about 78% accuracy



Correct Identification



Incorrect Identification



Workflow developed by Honggeun Jo, PhD student at The University of Texas at Austin.

Subsurface Data Analytics and Machine Learning

Machine Learning Examples

Lecture outline . . .

- Convolutional Neural Networks (CNN)

Instructor: Michael Pyrcz, the University of Texas at Austin

Convolutional Neural Nets

- A type of deep neural net used for image data, **2D information**
- Hierarchical approach segmenting image into smaller simpler patterns
- Inspired by visual cortex of animals
- **Convolution** step identifies features with trained filters
- **Subsampling** extracts features to a lower dimensional summary

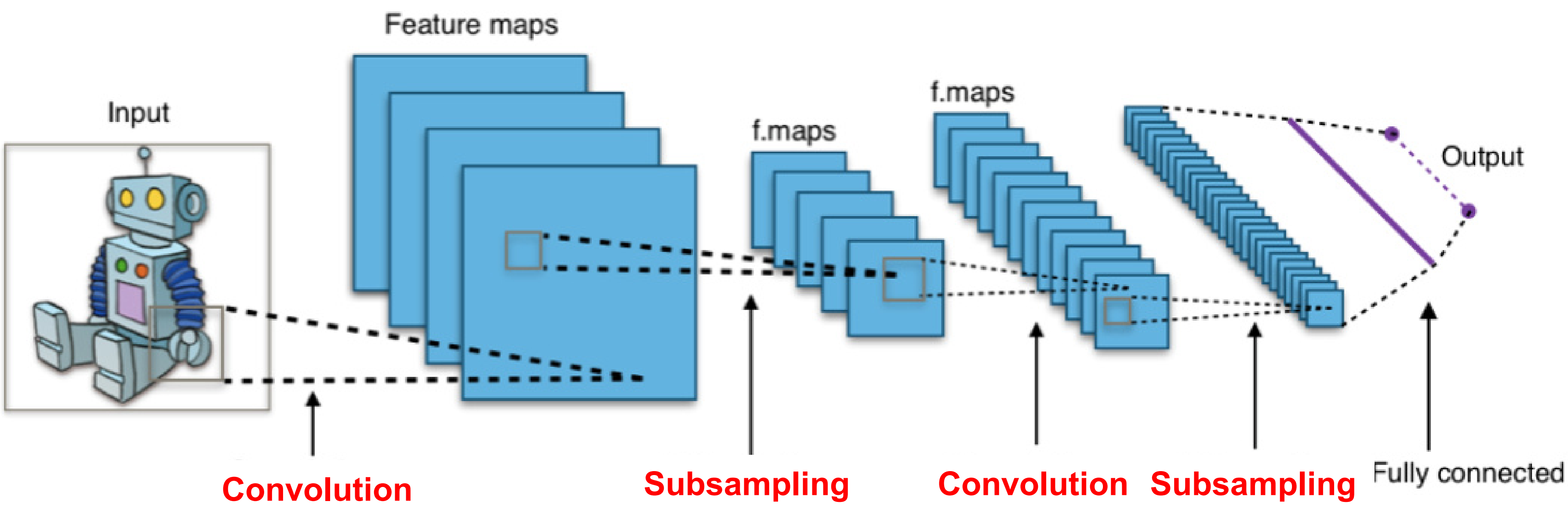
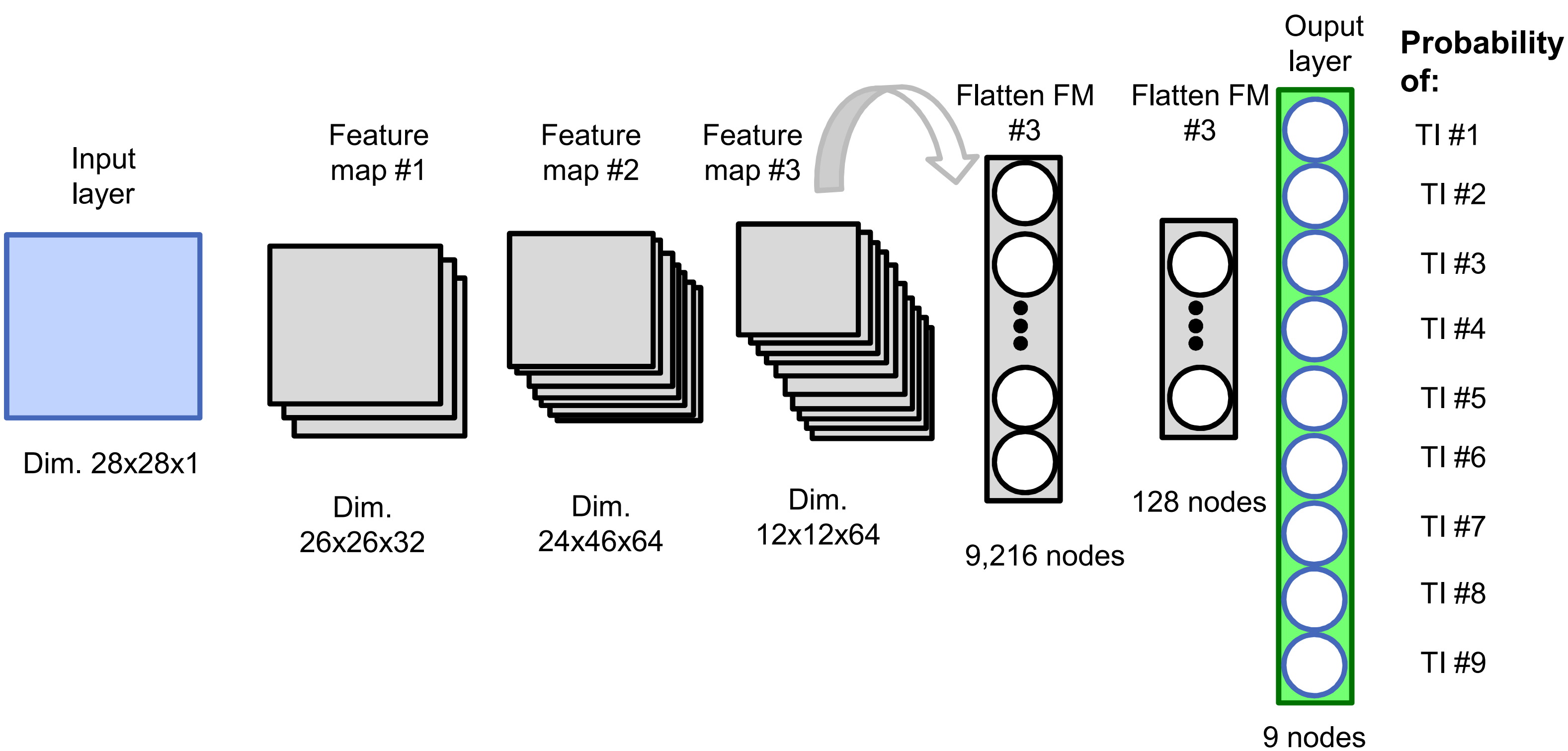


Image from https://en.wikipedia.org/wiki/Convolutional_neural_network#/media/File:Typical_cnn.png

Convolutional Neural Nets

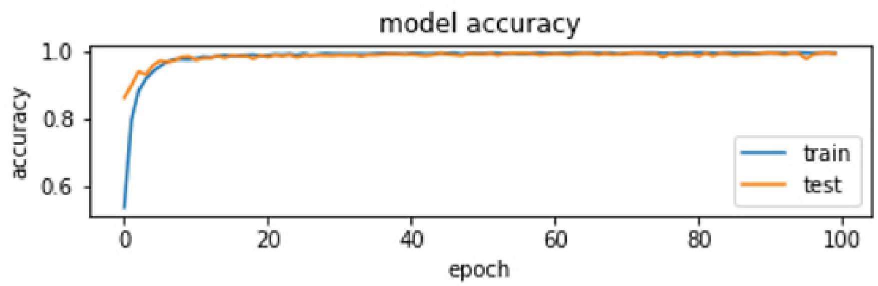
- Could we improve the previous result by moving to convolutional neural nets.
- Here's the design of a CNN to determine the probability of training image 1-9.



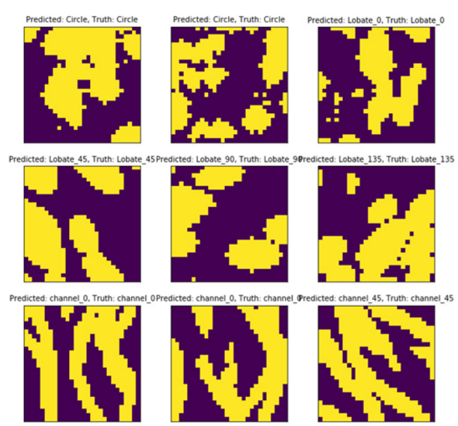
Workflow developed by Honggeun Jo, PhD student at The University of Texas at Austin.

Convolutional Neural Nets

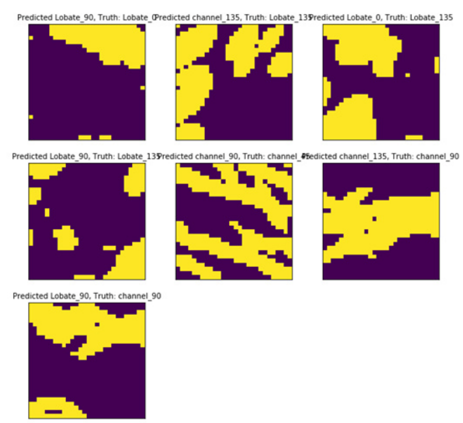
- The training and testing accuracy vs. number of training cycles, **Epoch**
- Achieved close to 100% accuracy



Correct Identification



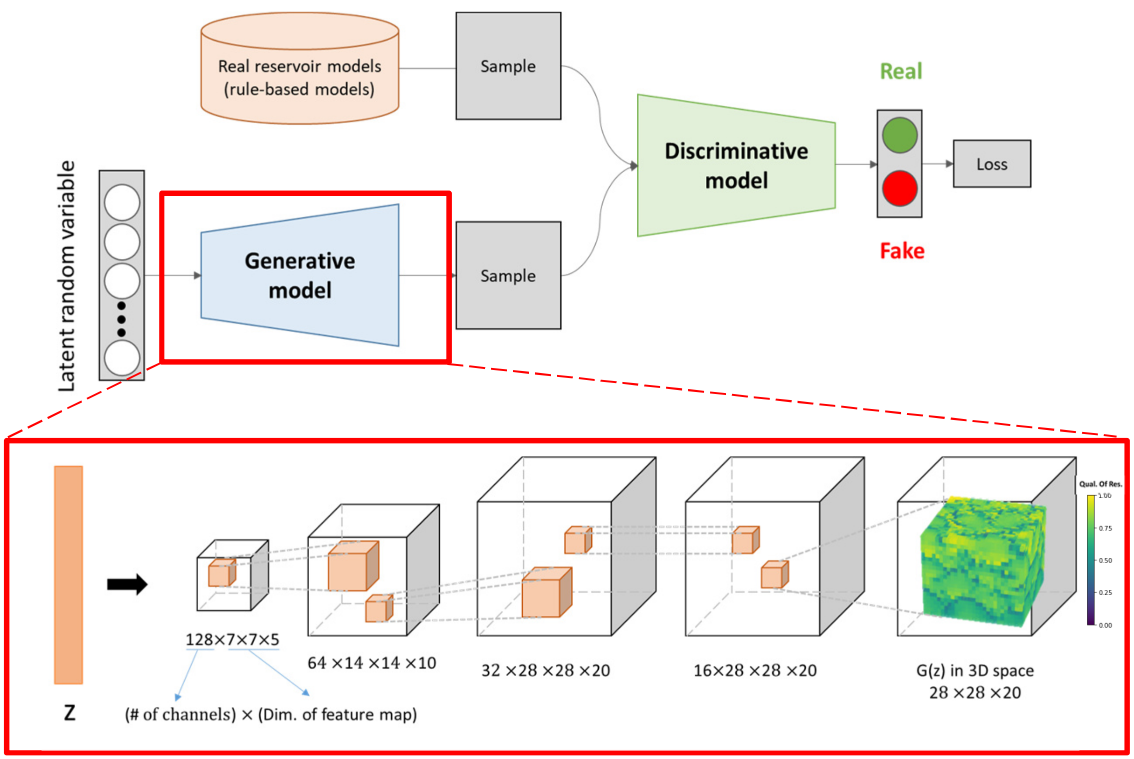
Incorrect Identification



Workflow developed by Honggeun Jo, PhD student at The University of Texas at Austin.

Convolutional Neural Nets

- Can we build reservoir models with convolutional neural nets?
  - We built a Deep Convolutional Generative Adversarial Network (DCGAN)
  - Generative model works to fool the discriminative model with fake models.



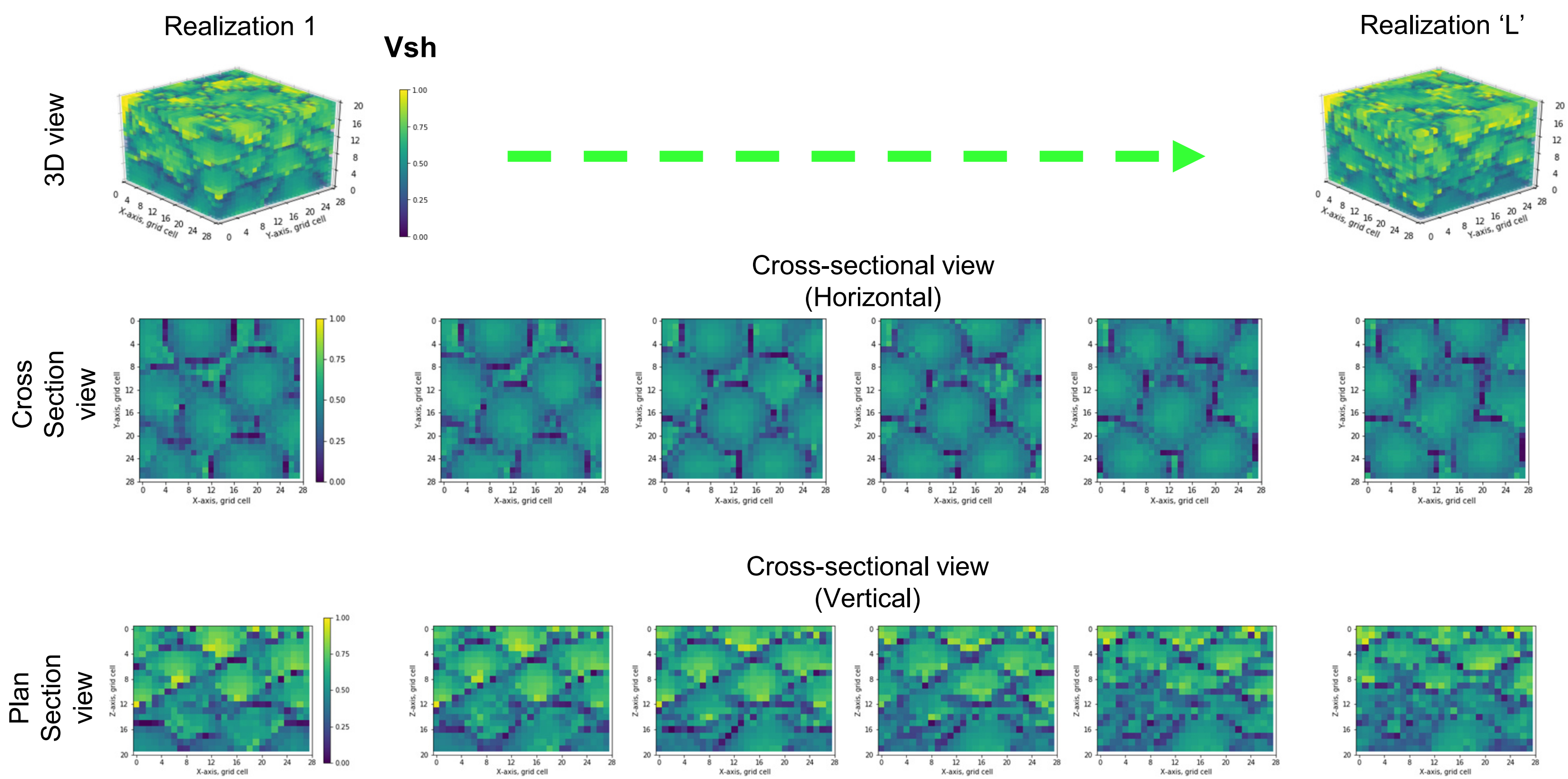
DCGAN workflow by Radford et al., 2015.

Workflow developed by Honggeun Jo, PhD student at The University of Texas at Austin.

Convolutional Neural Nets

Can explore the space of uncertainty along a continuous manifold.

- A latent reservoir manifold based on a single paramter



Workflow developed by Honggeun Jo, PhD student at The University of Texas at Austin.

Convolutional Neural Nets

Filling In Missing Spatial Information

- Semantic inpainting algorithm (Yeh et al., 2015).
- Using conceptual and perceptual information

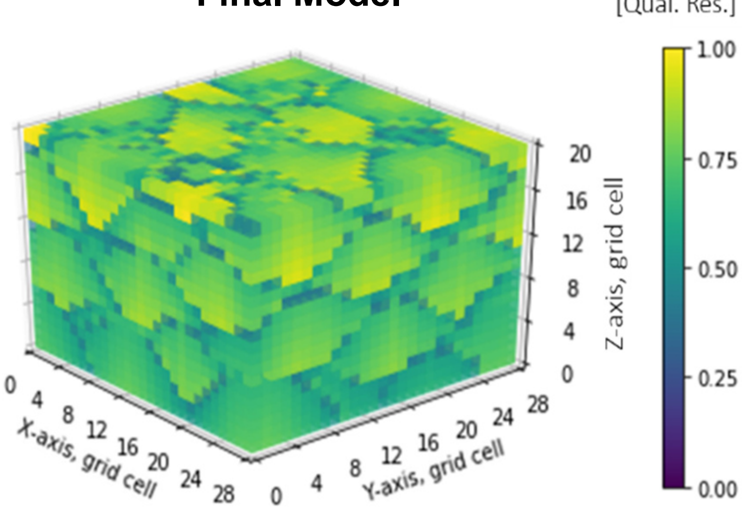
Examples of semantic image inpainting with DCGAN (Yeh et al., 2016)

Convolutional Neural Nets

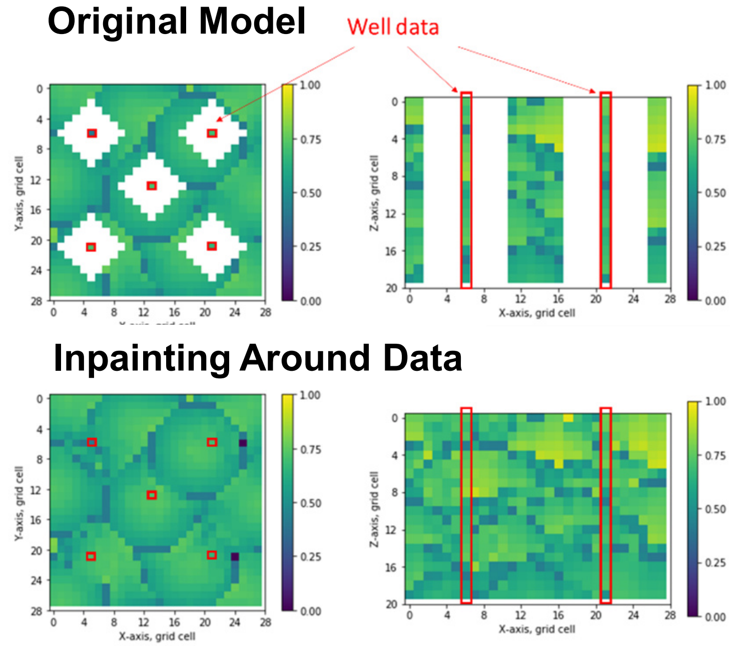
Conditioning to Well Data?

- Remove model around data
- Use conceptual (model around mask) and perceptual (model elsewhere to fill in missing model consistent with data

Final Model



Original Model



(b)

Workflow developed by Honggeun Jo, PhD student at The University of Texas at Austin.

Subsurface Data Analytics and Machine Learning

Machine Learning Examples

Lecture outline . . .

- Long Short-Term Memory Networks

Introduction

Data Analytics

Inferential Methods

Predictive Methods

Advanced Methods

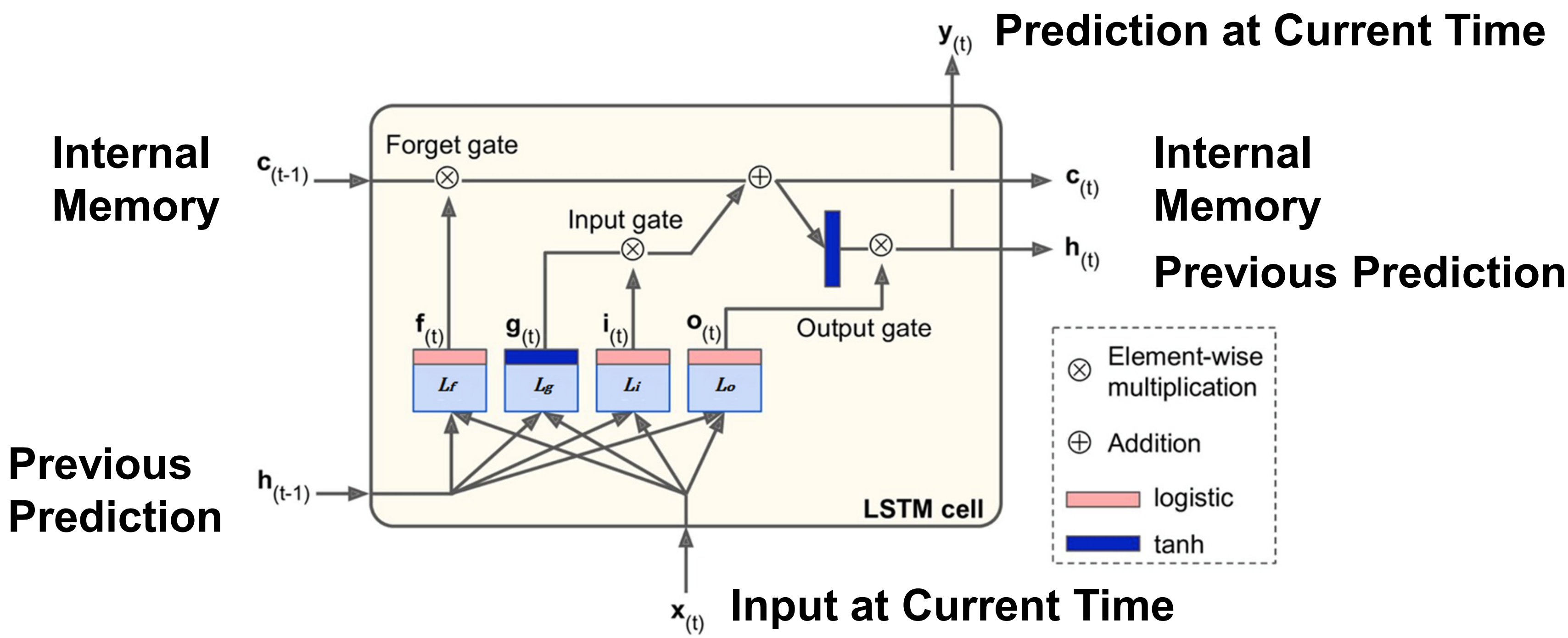
Conclusions

Instructor: Michael Pyrcz, the University of Texas at Austin

Long Short-Term Memory Networks

Neural networks often don’t perform well with time series data as they do not hold memory.

- The Long Short-Term Memory (LSTM) Networks approach combine previous long and short term experience.



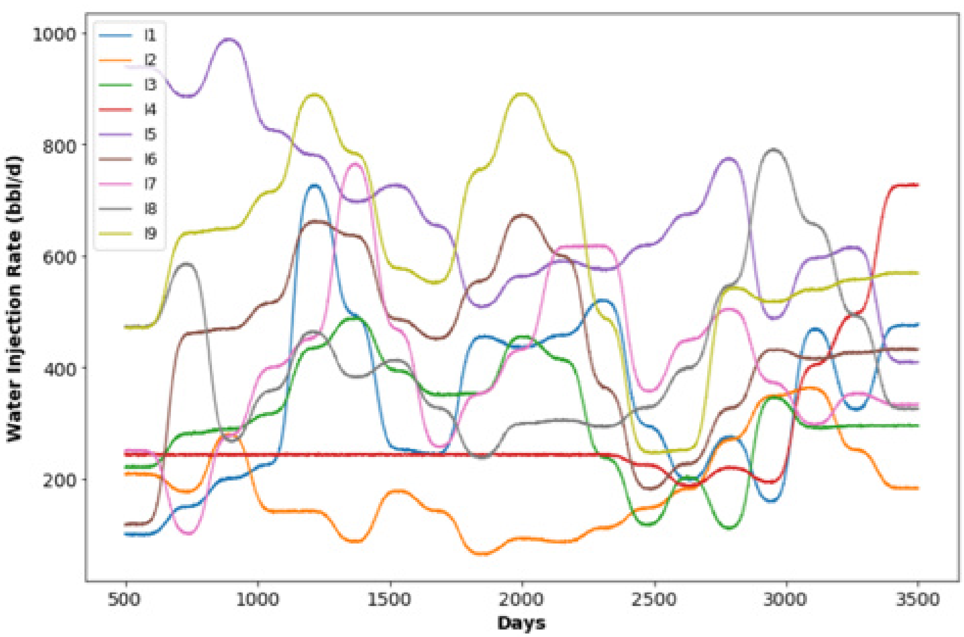
Workflow developed by Azor Nwachukwu, PhD student at The University of Texas at Austin, image from Géron, 2017 .

Long Short-Term Memory Networks

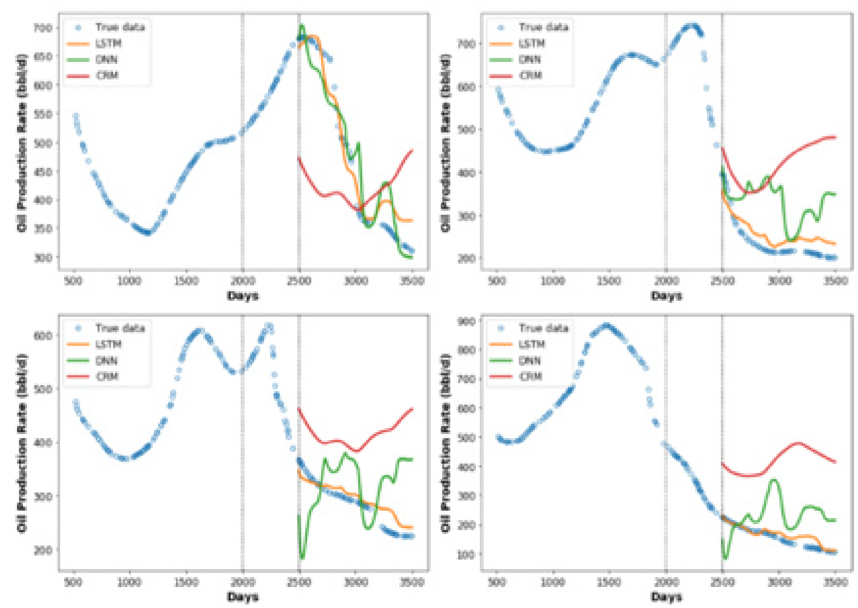
Prediction of producer flow rates based on complicated interactions of injectors.

- Train with 2500 days and predict future 100 days.

Injection Rates Over Train and Test Intervals



Production Over Train and Modeled Over Test



Workflow developed by Azor Nwachukwu, PhD student at The University of Texas at Austin.

Advanced Machine Learning Applications

Topic	Application to Subsurface Modeling
Support Vector Machines	Powerful method for developing segmentation decision rules. *Use to maximize differentiation in space, given labeled training data.*
Artificial Neural Nets	Flexible prediction models. *Use to formulate complicated decision rules, but loses spatial context, and very parameter rich.*
Convolutional Neural Nets	Flexible prediction models accounting for 2D hierarchical arrangements and features. *Use to explore, test and build models.*
Long Short-Term Memory Networks	Flexible prediction models accounting for 1D arrangements and features while accounting for memory. *Use to explore, test and build models predicting time series.*

Subsurface Data Analytics and Machine Learning

Machine Learning Examples

Lecture outline . . .

- Advanced Workflows
- Support Vector machine
- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- Long Short-Term Memory Networks (LSTM)

Introduction

Data Analytics

Inferential Methods

Predictive Methods

Advanced Methods

Conclusions

Instructor: Michael Pyrcz, the University of Texas at Austin