Intro to Machine Learning for Earth Scientists
SIOC 221B, Winter 2023
Lauren Hoffman
Lecture Notes

<table>
<tr><td>Slide No.</td><td>Topic / Main Takeaway</td></tr>
<tr><td colspan="2">A case study: Machine learning is a useful tool to predict and understand sea-ice motion in the Arctic.</td></tr>
<tr><td>4-9</td><td>Motivation: As sea-ice in the Arctic melts, machine learning is useful for…
Predictability: New opportunities for trade / transportation through the Arctic require knowledge and predictability of ice motion
Understanding: The relationship between ice-motion and wind is changing - can we use ML to understand how and why these changes are occurring?</td></tr>
<tr><td>10-11</td><td>ML vs. physics-based models:
ML is more computationally efficient than traditional physics-based models. Additionally, because it is not driven by physics, ML can also be used to draw information from the data and unravel emergent behaviors we have not yet recognized.</td></tr>
<tr><td>12-13</td><td>Choosing model inputs (predictors) and outputs:
We turn to the momentum balance for sea ice to determine what factors are important in predicting it’s motion
Present-day wind velocity, previous-day ice velocity, and previous-day ice concentration are used as inputs to predict present-day ice velocity
A combination of satellite products and reanalyses are used to train the ML models.</td></tr>
<tr><td>14</td><td>Model descriptions:
Three different models are applied to analyze the effectiveness of using a non-linear neural network (i.e. persistence vs. linear regression vs. neural network)</td></tr>
<tr><td>15-16</td><td>Evaluating overall model performance:
The model performance is evaluated AFTER the model has been trained on DATA THE MODEL HAS NOT SEEN.
The correlation (between actual data and what the model predicted) is used to evaluate the model.
The CNN outperforms the other models (i.e. has a higher correlation), both overall (slide 17) and spatially (slide 18, most regions in Figure 2 are red, indicating the correlation of the CNN is greater than that of the LR). This indicates that models that incorporate non-linear relationships between the inputs (i.e. the CNN) capture important information.</td></tr>
<tr><td>17-20</td><td>Using ML to understand physics with LR parameters & explainable AI:
The magnitude of the linear regression parameters (i.e. the A, B and C of the equation; the constants that are solved for by minimization of the cost function and</td></tr>
</table>

	then multiplied by the inputs) are indicative of how important each of the inputs was in predicting the output. ● Here, the parameter for wind speed, A, (Figure 1a) has the largest 'relevance' in predicting the ice velocity. The wind speed is particularly important in the central Arctic in comparison to coastal regions. ● We employ a technique known as 'Explainable AI' (XAI) to understand which inputs were relevant in predicting the output for the CNN (similar to looking at the LR parameters, but for a CNN). ● The XAI method used here (called Layer-wise relevance propagation) provides a mapped output called a "relevance heat map" for each gridpoint, time step, and input parameter. We will use this information to understand how the relationship between ice motion and wind is changing in time.
Introduction to ML	
21-27	There are many different types of ML models: this talk will focus on using **Neural Networks for regression problems.** ● There are many different machine learning methods / types of ML models. This talk will focus on neural networks. **We do not go into detail here about the different types of models, just show them for your reference.** ● There are also many different types of neural networks. Depending on your application it may be beneficial for you to choose one over the other. **We do not go into detail here about how to make those decisions here.**
What is the machine "learning" ?	
28-30	Neural network architectures consist of: ● **nodes** (also called neurons) ● **layers**: input, output, and hidden layers Information is passed between nodes depending on the NN parameters, namely the: ● **weights, w:** determine connections between nodes of different layers ● **biases, b:** how high the weighted sum ($x_i w_i$) needs to be for activation of a neuron ● **activations, a or f:** determine whether the information from a particular node is passed on to the next layer
31-33	Overview of linear regression: ● Linear regression is a type of machine learning that uses minimization of a cost function to solve for the model parameters. ● The cost function is related to the misfit, i.e. the difference between the model prediction and the actual data.
34-39	Overview of machine learning using neural networks ● Neural networks are also built on the basis of minimization of a cost function ● Neural networks work to minimize the cost function through an iterative process.

<table>
<tr><td></td><td><li style="list-style-type:circle">Each iteration is called an ‘epoch’.
<li style="list-style-type:circle">In each iteration the model parameters (weights, biases and activations of each node within each layer) are updated through a process called gradient descent. We do not discuss the details of gradient descent here.
<li style="list-style-type:circle">The cost function is evaluated during training of the model. Other metrics can be evaluated during the validation and testing phases, discussed below.</td></tr>
<tr><td>40-42</td><td>Setting up the data: before input into the model it is important to apply feature scaling and data splitting to your data:
Feature scaling makes it so that inputs are on similar scales, and can be applied in the form of either:
Standardization to zero mean and one standard deviation (subtract mean, divide by standard deviation)
Normalization to the maximum: (data - max) / (min - max)
Data splitting allows you to test your model on data it has never seen before. Data are divided into the following with an 80%-10%-10% split.
Training: the subset of data that the model uses to learn
Validation: data subset used to optimize the hyperparameters during training
Testing: data subset used to evaluate the model on data it hasn’t seen before</td></tr>
<tr><td>43</td><td>ML resources: where do you go to learn more about machine learning?</td></tr>
<tr><td colspan="2">Let’s Build a Neural Network</td></tr>
<tr><td>44-47</td><td>Setting up the problem for the tutorial.
Everything required to run this tutorial can be found at: https://github.com/lahoffman/ml_tutorial</td></tr>
</table>

Machine Learning Resources:

Steps I took to learn ML:

1. Cognitive Class
   Python: https://cognitiveclass.ai/courses/python-for-data-science
   AI: https://cognitiveclass.ai/courses/machine-learning-with-python

2. Coursera [free 1 week trial - you can finish the course in one week if you plan accordingly]
   https://www.coursera.org/learn/neural-networks-deep-learning

~~~~~~

Other Resources:

Beucler Lab at UNIL [data-driven atmospheric and water dynamics]
https://wp.unil.ch/dawn/getting-started-with-machine-learning/#site-header

Barnes lab colab notebooks:

CNN:
https://colab.research.google.com/github/eabarnes1010/course_ml_ats/blob/main/code/cnn_conus_clouds.ipynb#scrollTo=z7kbrJHQTY8X

ANN:
https://colab.research.google.com/github/eabarnes1010/course_ml_ats/blob/main/code/ann_basic.ipynb#scrollTo=EQk4PAhJIt7a

AI2ES Educational Resources
https://www.ai2es.org/products/education/

Datasets for ML:
http://mldata.pangeo.io/index.html

AI2ES/CIRA XAI Short Course:
https://docs.google.com/document/d/1lqpABwDl3kPe6ThE-NIDR64PimnltJEuKNkysDZuWKQ/edit

CIRA ML for Weather and Climate
https://docs.google.com/document/d/1SPNxZrbHMaIEaS2dbntDow9x_tgSuFTUTOugfa2NuRo/edit

Getting Started with NNs (non earth science)
https://github.com/fchollet/deep-learning-with-python-notebooks

ECMWF MOOC for MLWC (Machine learning for weather ad climate)
https://lms.ecmwf.int
~~~~~~