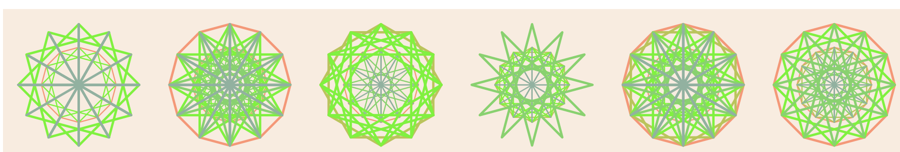
Review

CSC316 Machine Learning
Professor Leon Tabak

02 March 2022



Not every student studied each of these topics in depth, but in our reading and in hearing our classmates' presentations we all learned something about these topics:

- We familiarized ourselves with a checklist for machine learning projects.
  - for what purpose will our client use our model?
  - can we build on work that we or others have already done?
  - what are the potential and limitations of our model?
- We used mathematical terms and symbols and our knowledge of the operations that they denote to describe our models.
  - derivative of a function of a single variable
  - $\frac{\partial f(x,y,z)}{\partial x}$: partial derivative of a function of multiple variables
  - $\nabla f(x, y, z)$: gradient of a function (key to stochastic gradient descent)
  - $\vec{u} \cdot \vec{v}$ or $\mathbf{v}^T\mathbf{u}$: dot product of two vectors (we defined a linear model with the dot product of a parameter vector and a feature vector)

- product of two matrices (each element of the product is the dot product of a row in the first matrix and a column in the second matrix)
- $\mathbf{M}^T$: transpose of a matrix
- $\mathbf{M}^{-\mathbf{1}}$: inverse of a matrix (we did not learn how to construct an inverse, but learned the computational complexity and learned that Singular Value Decomposition gives us a means of constructing a pseudo-inverse)
- properties of the exponential function (e^{100} is a very large positive number, e^{-100} is a positive number very near 0.0)
- properties of the logarithm function ($\log 0$ is undefined, $\log 0.0001$ is a negative number with a very large magnitude, $\log 0.9999$ is a negative number whose value is just a little less than 1.0, $\log 1.0 = 0.0$)
- logistic (sigmoid) function

$$f(x) = \frac{1}{1 + e^{-x}}$$

- softmax (normalized exponential) function
- other activation functions—ReLu, tanh

- We used statistics to evaluate our models. Our algorithms used statistical measure to generate successively better guesses.
  - precision, recall, F1, ROC curves
  - root mean square error (RMSE) (also called Euclidean norm, ℓ_2 norm)
  - also, mean square error and mean absolute error
  - logistic regression cost function (log loss)
  - correlation matrix (Pearson's r)
  - confusion matrix
  - Gini impurity measure and entropy (for those who chose to study decision trees)
  - inertia and silhouette score (for those who chose to study clustering)
- We saw some widely datasets that many students and teachers use in their study of machine learning.
  - California Housing Prices dataset
  - MNIST dataset (hand-drawn numerals)
  - MNIST fashion dataset
  - Iris dataset

  - moons dataset

- We explored and prepared data. We used the Python programming language and the NumPy, Pandas, and Scikit-learn libraries. We used Matplotlib and Markdown in Jupyter notebooks to share our work.
  - divide dataset into training and test sets
  - examine first/last few instances
  - produce means, minima, maxima, standard deviation in columns
  - discard outliers
  - count nulls in rows and columns
  - drop rows/columns that contain a large number of nulls
  - replace nulls with mean, median, mode, or a constant value
  - scale data ($mean = 0, stddev = 1.0$ or $min = 0.0, max = 1.0$)
  - one hot encoding of categorical variables
  - engineer features (combining or dividing variables)
  - reduce the number of dimensions (although it was not part of the assigned reading, we identified PCA—Principal Component Analysis—as a means of reducing dimensions)
- We saw the tradeoff between bias and variance. We learned how to improve a model when we see evidence of underfitting and overfitting (for example, by regularization).
- We learned how to use cross-validation to develop our models. We learned how to search for optimal values for our hyperparameters by random assignments and by systematic search in a grid.
- We experimented with prediction, classification, and clustering.
  - linear regression
  - logistic regression
  - decision trees, random forests, support vector machines, k-means
  - neural networks (with Keras and TensorFlow)
    * specify number of layers
    * specify number of neurons in each layer
    * select activation functions
    * decide how layers will be connected
- We saw a variety of interesting applications of machine learning—in games, aviation, medical care, natural language processing, and the filtering of recordings of music and voices.
- We took just a little peek at some of the challenges in machine learning, looking, for example, at the amount of energy required to train some models.