DATASCI 207

Nedelina Teneva, PhD
nteneva@berkeley.edu

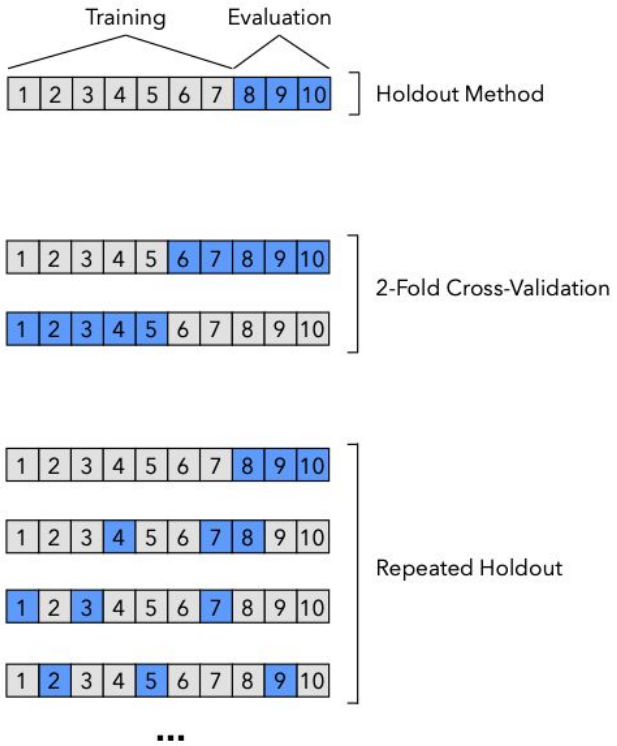
School of Information, UC Berkeley

Announcements

- **Finalize group selections in class**
  - Logistics Sheet
- **Evaluations:** https://course-evaluations.berkeley.edu/
  - What you like about the class + constructive feedback

Cross Validation (CV) Recap

CV



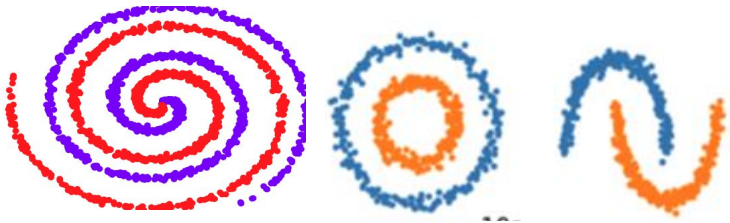
- Is there a difference between the hold out method with a 50/50 train/validation split and 2-fold cross validation?

- We can repeat the hold out methods multiple times (bottom panel). Would that be the same as doing k-fold cross validation?

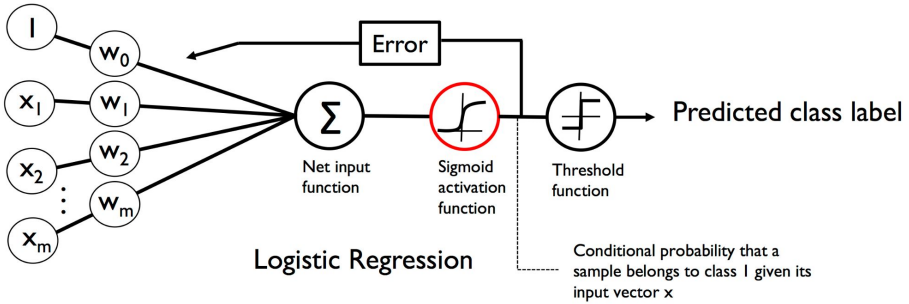
Multilayer Neural Networks

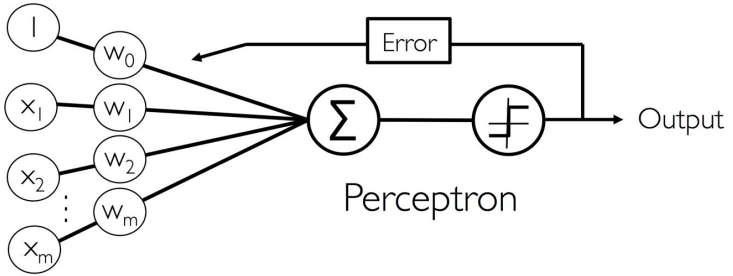
Feedforward Neural Networks (FNNs)

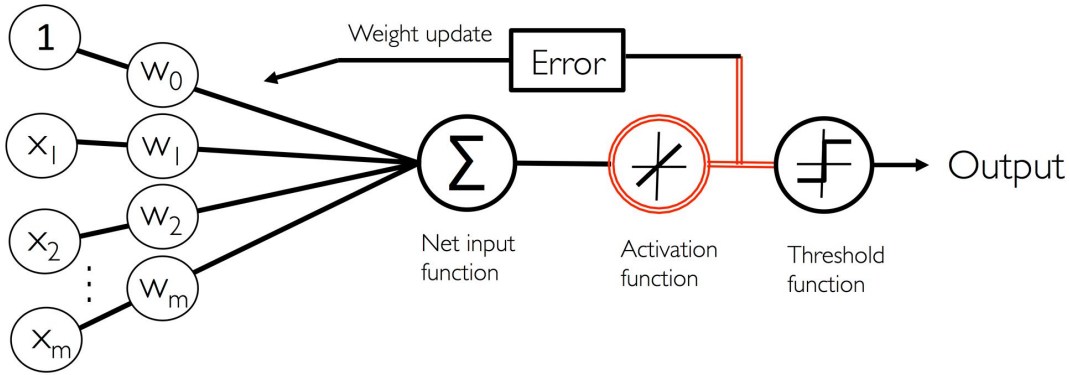
- Examples
  - Linear neural network - next slide
  - Single Layer Perceptron - last week's lecture
  - Multi Layer Perceptron (MLP) - XOR example from async
- Fully connected
- No cycles, unlike Recurrent Neural Networks (RNNs)



FNN Examples



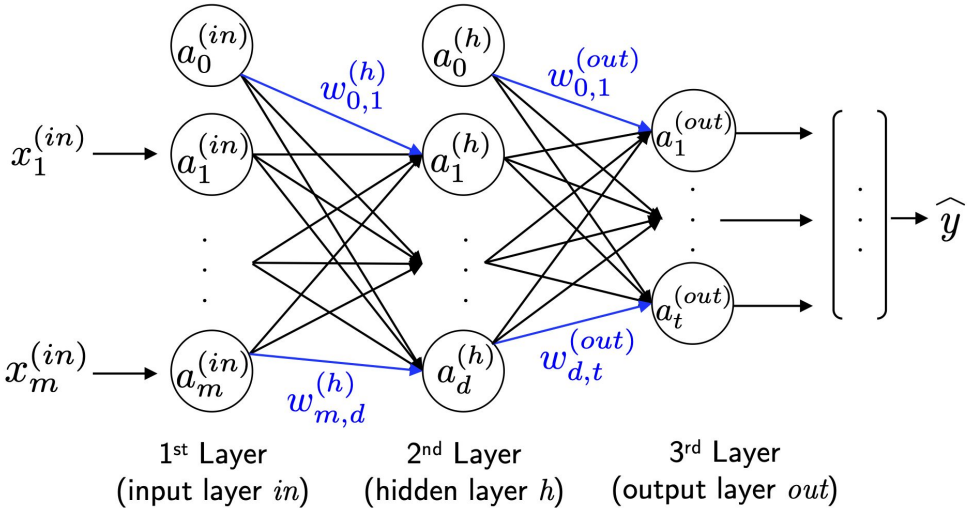




Source:
https://github.com/rasbt/machine-learning-book/blob/92e2320709071b4dcc82a99552aec145536cfbcd/ch02/ch02.ipynb

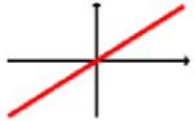
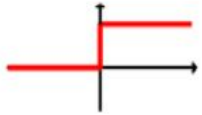
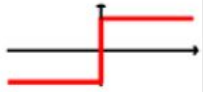
Multilayer Neural Networks

- (Right) Network with 1 input layer, 1 hidden layer, and 1 output layer. Each layer receives its inputs from the previous layer.
- We can add as many hidden layers (**deep neural networks**), but this becomes very computationally intensive (parameters to be estimated grow exponentially).
- **a()**: activation function
- Note that there are many activation functions to choose from (earlier we covered linear and sigmoid).
- Use nonlinear activation functions to account for the complexities of data.



Key Concepts

- **Key idea**: use of non-linearities (through the use of activation functions)
- **Optimization**: gradient descent, learning rate, different types of optimizers
- **Computational challenges**: computing the derivatives (forward pass and back propagation)
    - These challenges gave rise to frameworks such as tensorflow and pytorch which take care of the computation graphs and automatic differentiation behind the scenes

Activation function	Equation	Example	1D graph
Linear	$\phi(z) = z$	Adaline, linear regression	
Unit step (Heaviside function)	$\phi(z) = \begin{cases} 0 & z < 0 \\ 0.5 & z = 0 \\ 1 & z > 0 \end{cases}$	Perceptron variant	
Sign (signum)	$\phi(z) = \begin{cases} -1 & z < 0 \\ 0 & z = 0 \\ 1 & z > 0 \end{cases}$	Perceptron variant	
Piece-wise linear	$\phi(z) = \begin{cases} 0 & z \leq -½ \\ z + ½ & -½ \leq z \leq ½ \\ 1 & z \geq ½ \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, multilayer NN	
Hyperbolic tangent (tanh)	$\phi(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$	Multilayer NN, RNNs	
ReLU	$\phi(z) = \begin{cases} 0 & z < 0 \\ z & z > 0 \end{cases}$	Multilayer NN, CNNs	

Multilayer Neural Network Exercise

- Real Data (multi-class) - in class
  - https://github.com/MIDS-W207/nteneva/tree/main/live_sessions_current/week6

	Group 1	Group 2	Group 3
Question 1			
Question 2 (best performance)			

- Synthetic Example (binary, danish dataset) - try on your own!
  - https://github.com/MIDS-W207/coursework_2022/blob/main/Demos_Santerre/Berkley207-6.ipynb