Basic Terminology	
Feature	A piece of data that is used as input, like a pixel of an image, or numerical values like cost of an item.
Label	The goal or target that the model is trying to predict, like figuring out if a spam email is spam or not.
Example	An example is a piece of data that contains both the features and the label so the model can learn the target.
Epoch	An epoch is a complete iteration over the training data, like flipping through a deck of flash cards, can be tuned to adjust accuracy.
Weights	Weights are parameters/numeric values that represent the strength of the connection of the input and the target output. These can be adjusted for accuracy.
Bias	Just like weights, Bias' are just parameters/numeric values that represent overall activation of the neurons in the model.
Neuron	A neuron is just made up of weights, bias, and an activation function
Activation Function	An activation function is a mathematical function that combines the inputs and the bias to produce an output signal.
Loss	Loss is the difference from the predicted value, with the true value. The goal is to adjust parameters to reduce loss.
Labeled data	Labeled data contains the features and the classified label so the model learns from it.
Unlabeled data	Unlabeled data is unclassified data for unsupervised learning. Useful because labeled data is hard to come by or long to label.
Layer	A layer in a neural network is a group of neurons that process a set of input data, apply some processing to it, and produce an output signal that is passed on to the next layer or to the final output of the network.
Neural Network	A neural network is a ML model composed of interconnected layers of nodes, or neurons, that process input data and produce output predictions. The neural network learns from data and adjusts its internal parameters to improve the accuracy of its predictions over time.
Model	A model is a representation of a system or a process that is created by training a machine learning algorithm on a dataset.
Training	Training is the process of teaching a model to recognize patterns in data and make accurate predictions. This happens by showing it labeled examples.

Basic Terminology (cont)	
Inference	Inference is the process of using a trained model to make predictions on new, unseen data.
Regression	Regression is a way of making predictions based on input data. For example, if you want to predict how much a house will sell for based on its size, location, and other features, you can use regression. The goal of regression is to find a mathematical formula that accurately predicts the output value (in this case, the sale price) based on the input values (in this case, the size, location, and other features of the house).
Classification	Classification is like sorting objects into different boxes based on their features. For example, if you want to sort fruit into boxes based on their color, you would put all the red apples in one box, all the green apples in another box, and so on. In machine learning, classification is a similar process, but instead of fruit, we sort data into categories or classes based on input features.
Hyper-Parameters	Hyperparameters are parameters that are set before the training of a machine learning model and determine the behavior of the training algorithm.
Gradient Descent	Gradient Descent is an optimization algorithm commonly used to find the optimal values of the parameters of a model that minimize a given cost or loss function.
SDG	Stochastic Gradient Descent (SDG) is a variant of gradient descent that updates the parameters of the model for each data point in the dataset. SDG is commonly used in machine learning for large datasets because it is faster and more efficient than batch gradient descent.
Batch Descent	Batch Gradient descent updates the model parameters using the gradient of the loss function computed over the entire training dataset.
Mini Batch Descent	Mini Batch Gradient Descent updates the model parameters using the gradient of the loss function computed over a small randomly selected subset of the training dataset.
Learning Rate	The learning rate is a hyperparameter that controls how much the model parameters are updated during training. It determines the step size at each iteration of the optimization algorithm, such as gradient descent or its variants.
Convergence	Convergence refers to the point at which the optimization algorithm has found the optimal values of the model parameters and has reached the minimum value of the loss function.

By **Water** (watersippin)

cheatography.com/watersippin/

Not published yet.
Last updated 20th April, 2023.
Page 2 of 3.

Basic Terminology (cont)	
Emperical Risk Minimization	Empirical Risk Minimization is a principle in machine learning and statistics that states that the best model for a given problem is the one that minimizes the empirical risk or training error.

Activations	
ReLU	The ReLU activation function is used to determine whether a neuron in a neural network should "fire" or not. When the input to the ReLU function is positive, the function outputs the input value. When the input is negative, the function outputs zero.
Sigmoid	Sigmoid is an activation function that maps any input value to a value between 0 and 1, and is commonly used in binary classification problems.
Softmax	Softmax is an activation function that is used to convert a vector of numbers into a probability distribution, and is commonly used in multiclass classification problems.

Algorithms	
MSE	Mean Squared Error (MSE) is a metric used to measure the performance of regression models. It calculates the average squared difference between the predicted values and the actual values for a set of data points. A smaller value of MSE indicates that the model is making more accurate predictions, while a larger value of MSE indicates that the model is making less accurate predictions.

Algorithms	
MSE	Mean Squared Error (MSE) is a metric used to measure the performance of regression models. It calculates the average squared difference between the predicted values and the actual values for a set of data points. A smaller value of MSE indicates that the model is making more accurate predictions, while a larger value of MSE indicates that the model is making less accurate predictions.