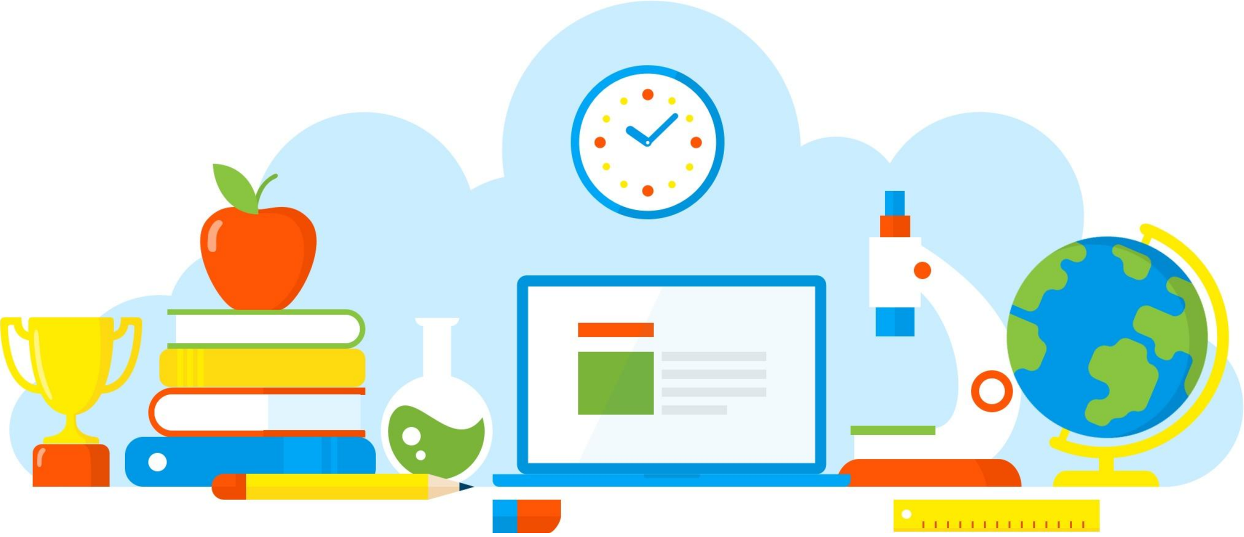


Let’s Go Python!

GBUS-401
Python Lab 2: Machine Learning Fundamentals

Zirong Chen
03/16/2021

Introduction

To me, Machine Learning can be understood as a process that **humans teach computers to learn something**, or humans make computers to **learn some general patterns in some tasks**.

There are several general types of Machine Learning Algorithms: **Unsupervised Learning**, **Supervised Learning**, **Reinforcement Learning**.

Supervised Learning. Given sample X and label y, figure out the function f, which is able to give the best f(x) that is closest to y. There are two types of Supervised Learning: Classification and Regression. Some terminologies:

- *Training set*(s): the dataset used to train, with X and y.
- *Validation set*(s): the dataset used to validate the model performance and normally is split from the training set, with X and y.
- *Test set*(s): the dataset is used to test the model performance, only with X.
- *Label*: usually refers to the y in the training set.
- *Prediction*: usually refers to the output from our trained model or y_hat.
- *Loss function*: one function used to evaluate the performance of our models.

Unsupervised Learning. Given sample X and to discover the potential relationships within the dataset. Most commonly used Unsupervised Learning Algorithm: Clustering.

Reinforcement Learning. Given an environment, take sequential actions to make the most rewards.

The First Step: Installation

Scikit-learn is one of the most popular packages for Statistical Machine Learning. It is open-sourced and easy to use. To install sci-kit learn, please refer to this website.

Keras, aka Tensorflow 2.x, is one of the most popular frameworks for Deep Learning. And it is also very user-friendly and open-sourced. To install Keras/Tensorflow, please refer to this website.

Supervised Learning Fundamentals

As mentioned above, there are two general types of Supervised Learning: **Classification** and **Regression**. Classification is to tell which category the input sample belongs to. Regression is to tell an actual value. For example, to tell whether a house is gonna be sold or not is a classification task; to tell how much will this house be sold is a regression task.

Loss function: for classification, usually is Cross-entropy; for regression, usually is Mean Squared Error. For example:

Houses	Groud Truth	Prediction
House1	1	0.3
House2	1	0.8
House3	0	0.1
House4	0	0.9

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Binary Cross-Entropy / Log Loss

Houses	Groud Truth	Prediction
House1	16,000	12,000
House2	152,000	155,000
House3	30,000	32,000
House4	17,100	15,000

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Evaluation methods: F1 score, confusion matrix, AUC/ROC…

Cross-validation: to split the dataset into several pieces, just in order to (1)better evaluate the model performance with a hand-crafted validation set; (2)pick up your best hyper-parameters for this task.

Statistical Learning Fundamentals(Linear Regression)

Basically, y_hat = WX + B. To figure out the best W and B that could give us the y_hat that is closest to y. How? To minimize the sum of distances from all data points to the line we draw.

Regularization: constrain the weights to avoid overfitting. Linear regression with L1 regularization: LASSO regression; Linear regression with L2 regularization: Ridge Regression.

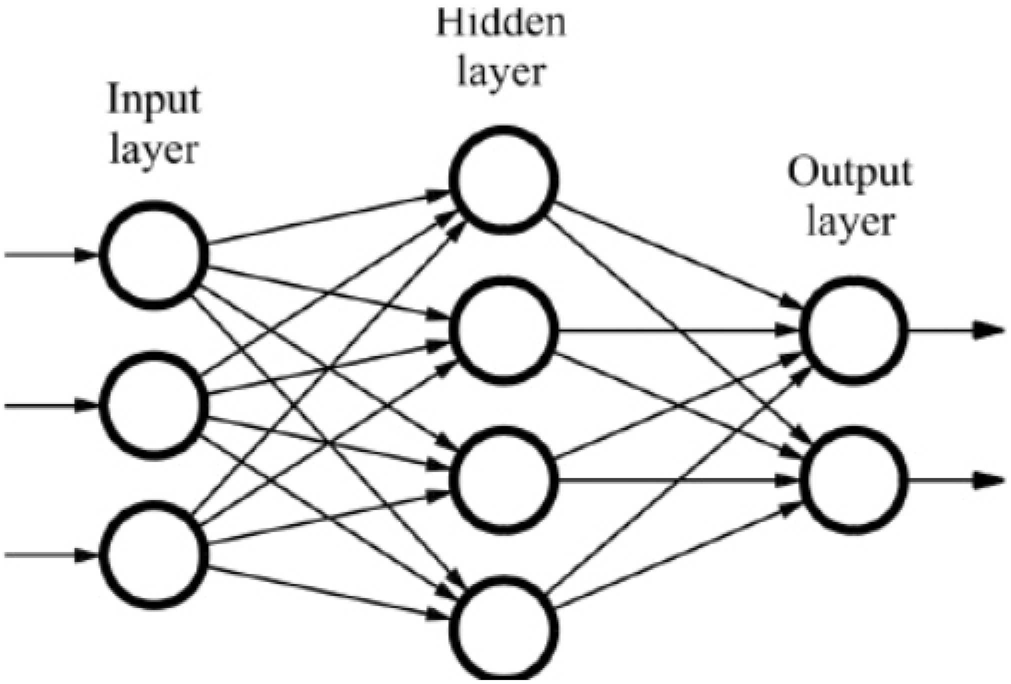
L1 Regularization

$$\text{Cost} = \sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij}W_j)^2 + \lambda \sum_{j=0}^{M} |W_j|$$

L2 Regularization

$$\text{Cost} = \underbrace{\sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij}W_j)^2}_{\text{Loss function}} + \underbrace{\lambda \sum_{j=0}^{M} W_j^2}_{\text{Regularization Term}}$$

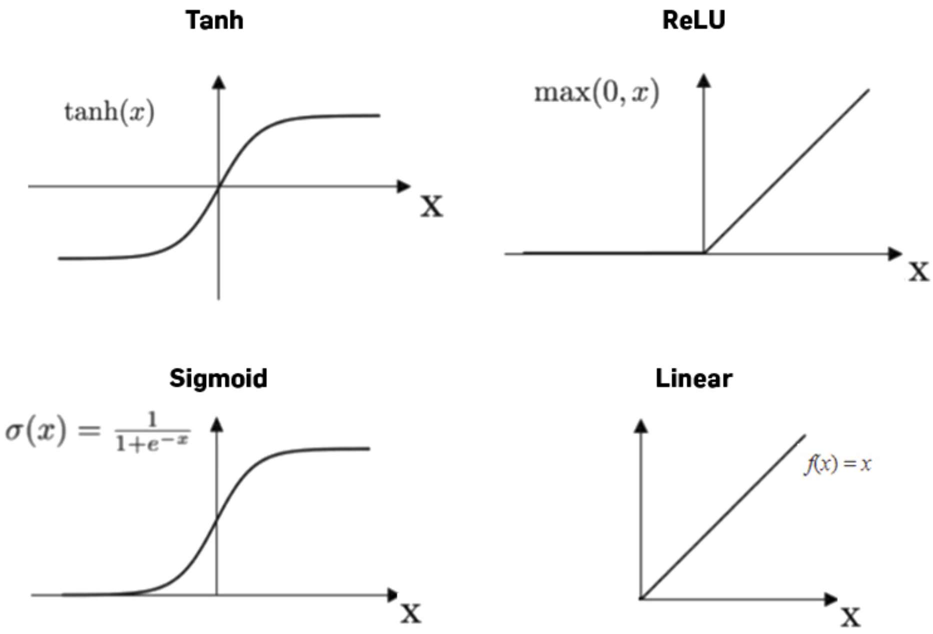
Deep Learning Fundamentals



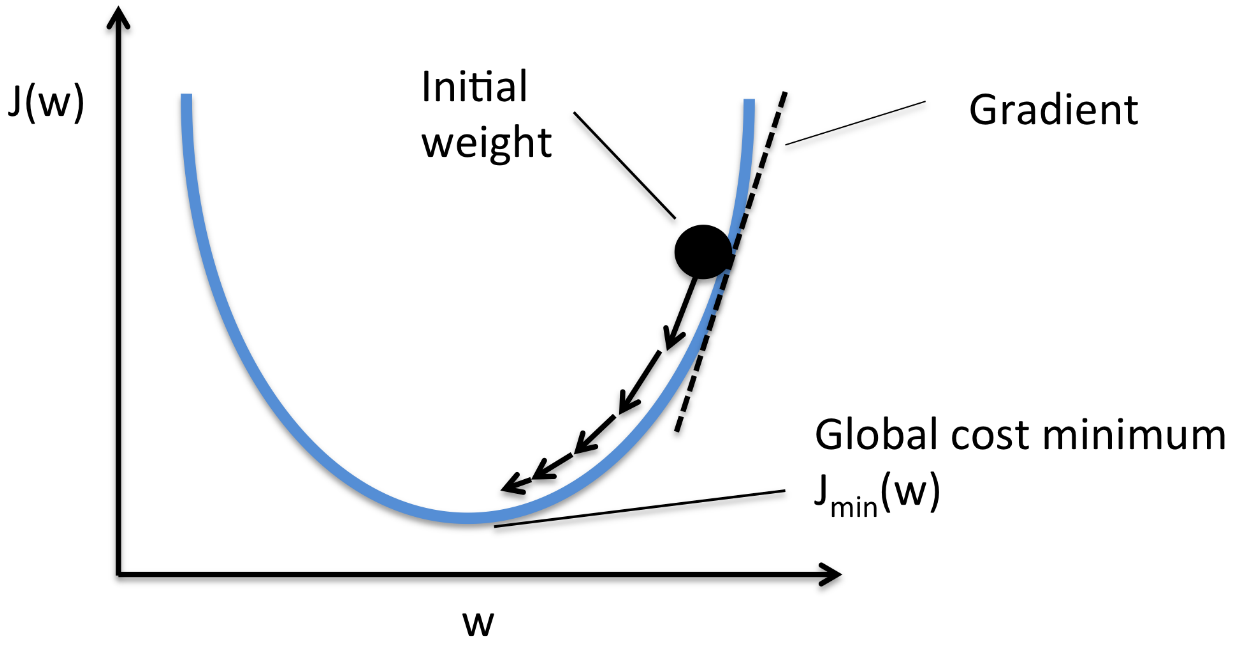
Terminologies:

- *Neuron*: nodes
- *Hidden layer*: see the picture above
- *Weights, Bias*: similar as the W and B in WX + B
- *Epoch*: learning iterations
- *Learning rate*: control the gradient descent
- *Optimizer*: built-in optimizers for gradient descent.

Activation functions: Sigmoid, ReLU, Tanh...



Gradient Descent: To iteratively get the minimum by derivatives from the loss function. The learning rate is used to control the “step size”.


J(w)
Initial weight
Gradient
Global cost minimum
J_min(w)
w