

Seunghyuk Baek

sbaek44@vt.edu

CS 5644 Fall, 2018

Homework 4

For this assignment, the popularity of a web site is predicted with MLP model in sklearn module in python. There are 8 possible neural network models.

1-1: Use 2 hidden layers with 20 nodes in each layer and use 'relu' activation function.

1-2: Use 2 hidden layers with 20 nodes in each layer and use 'tanh' activation function.

2-1: Use 2 hidden layers with 100 nodes in each layer and use 'relu' activation function.

2-2: Use 2 hidden layers with 100 nodes in each layer and use 'tanh' activation function.

3-1: Use 5 hidden layers with 20 nodes in each layer and use 'relu' activation function.

3-2: Use 5 hidden layers with 20 nodes in each layer and use 'tanh' activation function.

4-1: Use 5 hidden with 100 nodes in each layer and use 'relu' activation function.

4-2: Use 5 hidden with 100 nodes in each layer and use 'tanh' activation function.

All models were cross validated with 5-fold cross validation; resulting precision, f1 score, and recall value.

The table below is summary of findings, L means hidden layers, N denotes the number of nodes per layer and relu/tanh represent activation function used.

		2L,20N,relu	2L,20N,tanh	2L,100N,relu	2L,100N,tanh	5L,20N,relu	5L,20N,tanh	5L,100N,relu	5L,100N,tanh
Precision	Average	0.55	0.54	0.55	0.57	0.52	0.50	0.51	0.58
	Standard Dev	0.08	0.06	0.10	0.02	0.08	0.09	0.09	0.08
recall	Average	0.75	0.21	0.70	0.61	0.52	0.51	0.68	0.47
	Standard Dev	0.36	0.10	0.38	0.33	0.40	0.36	0.41	0.44
f1 score	Average	0.38	0.39	0.28	0.37	0.54	0.38	0.27	0.44
	Standard Dev	0.26	0.23	0.20	0.15	0.26	0.30	0.23	0.24

First, lets look at the effect of activation function. Generally, the effect of changing activation functions is reduction of the average values of precision, recall and f1. However, the standard deviation of measured values also tends to decrease when the activation function is changed from 'relu' to 'tanh'. The effect of the number of nodes was the greatest on the recall value of 2-layer architecture with hyperbolic tangent activation function. Changing the number of hidden layers effect the result when the number of nodes is relatively small (20 nodes). All of architectures show relatively stable precision values with small standard deviation. However, recall values vary from 21 to 75 and standard deviations are much higher than that of average values. In addition, average f1-scores are very low compare to precision and recall averages. This can be an indicator for poor prediction on positive testing result

creating too many false negatives. To pick a best choice, the architecture with 5-layers and 20 nodes with 'relu' activation function seems the best choice since its precision is very similar to others and has the best f1-score.

Task 2

		Learning Rate 0.001	Learning Rate 0.5
Precision	Average	0.51	0.16
	Std	0.09	0.20
recall	Average	0.68	0.60
	Std	0.41	0.49
f1 score	Average	0.27	0.24
	Std	0.23	0.30

The learning rate in neural network is a constant that changes the weight of features. In every iteration, machine learning process adjust weight of each features so that the error between predicted value and the real value eventually become minimum. Higher learning rate means the machine learning process will change weight more dramatically to achieve minimum error. For this task, I compared precision, recall and f1 values of 5-fold cross validation using 5-layer, 100 nodes and 'relu' activation function. As result, higher learning rate decreases all these values because the machine learning process judged weights of each features too quickly.