CS-6350: HW 2

James Brissette

September 28, 2018

1 Linear Classifiers and Boolean Functions

1.

2.

3.

4.

5.

2 Feature transformations

1.

2.

3 Mistake Bound Model of Learning

1.

   (a)

   (b)

2. A linear classifier compatible with the given data set is given by the below

   (a)

   (b)

4 The Perceptron Algorithm and its Variants

1.

2.

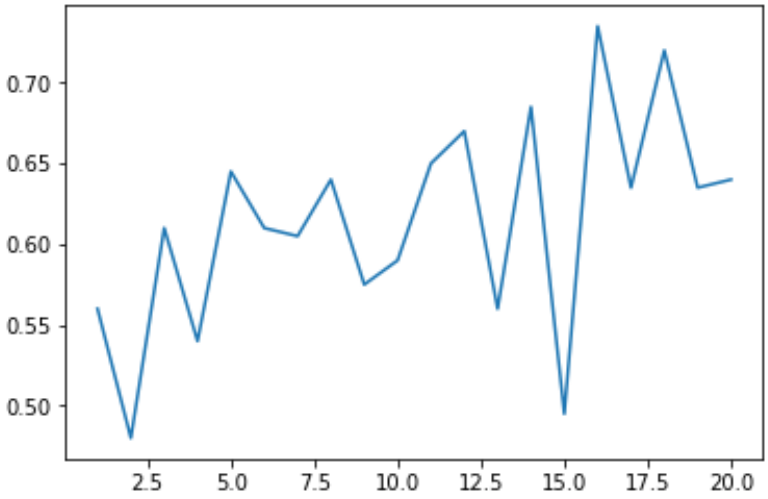
3. Noting that the random initialization may sometimes give different results with respect to the best hyper-parameters, in most cases I ran the 5-fold cross validation (for ten epochs each) several times and averaged them to see, on average, which parameters were actually the best. My results were as follows:

   Simple Perceptron
   a) rate = {.01}

b) cross-validation accuracy using rate=.01 was **0.612733**
c) total number of updates on the training set was **6015**
d) the best accuracy on the development set came with using 16 epochs and yielded an accuracy of **.735** (see below for the plot used to determine the best epoch to use). This yielded a weight vector, bias combination as follows:

$$w = [-4.790801, -6.310801, 18.629199, 0.809199, -9.290801, \quad (1)$$
$$-8.960801, -3.480801, 1.899199, 0.75067916, -1.05087573, \quad (2)$$
$$0.91483852, 1.0267039, 3.30614975, 1.8007971, 0.88184558, \quad (3)$$
$$0.41990239, -2.7867524, -0.66128905, -2.170801] \quad (4)$$
$$b = -4.567802999999947 \quad (5)$$

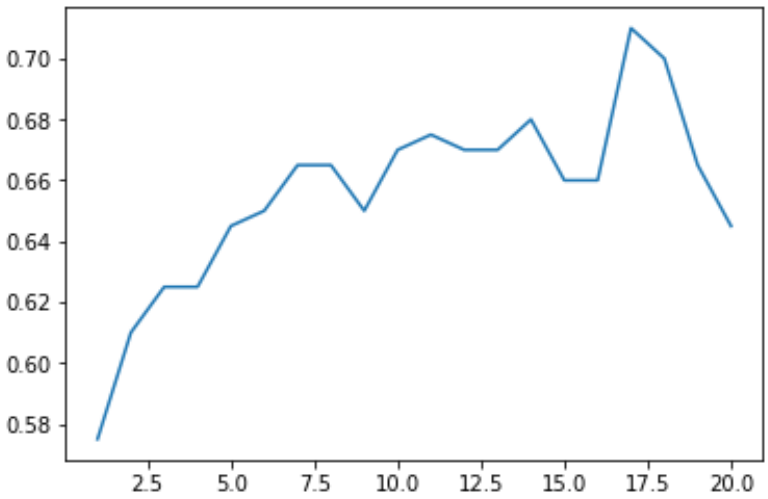
e) using the weight vector and bias from epoch 16 to predict on the test set, the algorithm reported an accuracy of **.587065**
f)



Decaying Perceptron
a) For the specific combination of the random shuffling of the training data and the initial weight vector and bias, I found the optimal initial rate = {1}
b) cross-validation accuracy using rate=1 was **0.664533**
c) total number of updates on the training set was **5113**
d) the best accuracy on the development set came with using 17 epochs and yielded an accuracy of **.71** (see below for the plot used to determine the best epoch to use). This yielded a weight vector, bias combination as follows:

$$w = [-0.52830582, -0.62562476, 4.58366762, 1.83460061, -0.48774808, \quad (6)$$
$$-2.43257835, -2.71244112, -1.11330101, 0.85833871, -1.40136228, \quad (7)$$
$$-2.3033998, -1.70711248, 0.57267534, 0.43230634, 0.20699507, \quad (8)$$
$$0.06603351, -0.31253566, -0.08022425, -0.59702606] \quad (9)$$
$$b = -32.005986 \quad (10)$$

e) using the weight vector and bias from epoch 17 to predict on the test set, the algorithm reported an accuracy of **.71144**
f)



Margin Perceptron

a) For the specific combination of the random shuffling of the training data and the initial weight vector and bias, I found the optimal combination of hyper-parameters to be margin=.01, rate=1

b) cross-validation accuracy using margin=.01, rate=1 was **0.6880**

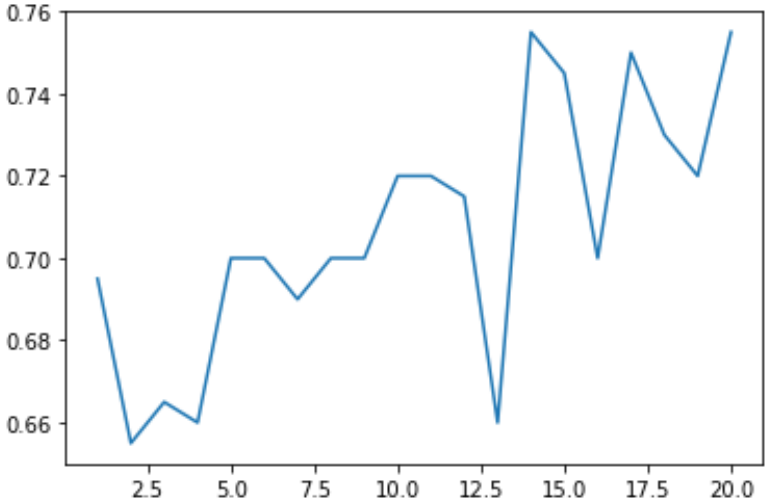
c) total number of updates on the training set was **4930**

d) the best accuracy on the development set came with using 14 epochs and yielded an accuracy of **.6** (see below for the plot used to determine the best epoch to use). This yielded a weight vector, bias combination as follows:

$$w = [-0.54624921, -0.95548707, 2.26801139, 0.52011042, -1.45902761, \quad (11)$$
$$-1.02471788, 0.10777735, -0.68208914, -0.01616318, -0.57514031, \quad (12)$$
$$0.97513725, 0.30675042, 0.21458043, 0.21317788, 0.17914802, \quad (13)$$
$$0.07836919, -0.30403893, -0.05202941, 0.01070719] \quad (14)$$
$$b = -1.008271999999998 \quad (15)$$

e) using the weight vector and bias from epoch 14 to predict on the test set, the algorithm reported an accuracy of **.721393**

f)



4.