CS 696(Machine Learning)

Assignment 3

Diwas Sharma
September 2017

1. You are given 21 well distributed training pattern of dimensionality 3.

a. What is the probability that a linear classifier separates a randomly selected dichotomy of training patterns?

We have N = 21 and n = 3, so the probability that a linear classifier separates a randomly selected dichotomy of training patterns is given by

$$\frac{D(N,n)}{2^N}$$

Where D(N,n) is given by,

$$D(N,n) = 2\sum_{k=0}^{n}\frac{(N-1)!}{(N-1-k)!\,K!}$$
$$= 2 + 40 + 380 + 2280$$
$$= 2702$$

So the probability is,

$$\frac{2702}{2^{21}}$$
$$= 0.0012$$

b. What is the minimum degree polynomial function that guarantees separation of any dichotomy of 21 sample patterns?

In the transformed space the pattern dimension of linear decision boundary must be greater than

$$\frac{N}{2} - 1$$

$$\geq \frac{21}{2} - 1$$

$$\geq 9.5$$

So the non linear decision boundary in original dimension must have at least 10 terms.

$$\frac{(3+r)!}{3!\,r!} \geq 10$$

$$r^3 + 6r^2 + 11r + 6 - 60 \geq 0$$

A solution for above inequality is given by $r \geq 2.00$

Thus, a second degree polynomial function in original sample dimension guarantees separation of any dichotomy.

2. You are given training patterns {X_1 , , ... , X_N } from both classes. All patterns are augmented by 1 and patterns of Class-2 are multiplied by -1 as discussed in the class.

The cost function J(θ) = ∑ I = 1 to N (1/X i t X i) [|θ t X i | 2 - |θ t X i |θ t X i] attains its only minimum value of zero if θ t X i > 0 for all X i .Develop a machine learning algorithm

based on gradient descent approach to learn θ from training data to separate the two classes.

The gradient of the cost function J(θ) wrt θ_j is,

$$\frac{\delta J(\theta)}{\delta \theta_j} = 2 \frac{Xij}{X_i^t X} \left[\frac{|\theta^t X_i|^2 - |\theta^t X_i|}{\theta^t X_i} \right]$$

So an algorithm based gradient descent approach to learn θ is as follows,

Step 1: Randomly initialize θ

Step 2: At each iteration i, present X_i

If $\theta(i)^t X_i > 0$ then

$$\theta(i+1) = \theta(i)$$

else

$$\theta(i+1) = \theta(i) + \alpha \times Xi \times \frac{1}{X_i^t X} \left[\frac{|\theta^t X_i|^2 - |\theta^t X_i|}{\theta^t X_i} \right]$$

$$i = i + 1 mod\ N$$

Step 3: If all parameters are correctly classified in a row then stop. Otherwise goto 2.

3. You are given the following training data.

Class1: $(2\ 4)^t$, $(3\ 3)^t$

Class2: $(6\ 12)^t$, $(8\ 10)^t$

Starting with an initial parameter vector $[0\ 1\ 1]^t$,

a.Illustrate 4 iterations of perceptron - formulation 1

At first we augment all the samples by 1 to obtain following

Class1: $(1\ 2\ 4)^t$, $(1\ 3\ 3)^t$

Class2: $(1\ 6\ 12)^t$, $(1\ 8\ 10)^t$

For iteration 1,

$$\theta^t X(k) = [0\ 1\ 1][1\ 2\ 4]^t = 6$$

Since the value is positive and data belongs to class 1 an update is not carried out, so

$$\theta^t(k+1) = \theta^t(k) = [0\ 1\ 1]$$

For iteration 2,

$$\theta^t X(k) = [0\ 1\ 1][1\ 3\ 3]^t = 6$$

Since the value this data is positive and belongs to class 1 as well an update is not carried out. So

$$\theta^t(k+1) = \theta^t(k) = [0\ 1\ 1]$$

For iteration 3,

$$\theta^t X(k) = [0\ 1\ 1][1\ 6\ 12]^t = 18$$

Since the value is positive but the data belongs to class 2 the following update is carried out,

$$\theta^t(k+1) = \theta^t(k) + \alpha X(k) = [0\ 1\ 1] + 1 * [1\ 6\ 12]^t = [-1\ -5\ -11]$$

For iteration 4,

$\theta^t X(k) = [-1 \ -5 \ -11][1 \ 8 \ 10]^t = -161$

Since the value is negative and data belongs to class 2 an update is not carried out. So,

$\theta^t(k+1) = \theta^t(k) = [-1 \ -5 \ -11]$

b. Illustrate 4 iterations of perceptron - formulation 2

At first we augment all the samples by 1 and multiply samples from class 2 by -1 to obtain following,

Class1: $(1\ 2\ 4)^t$, $(1\ 3\ 3)^t$

Class2: $(-1\ -6\ -12)^t$, $(-1\ -8\ -10)^t$

For iteration 1,

$\theta^t X(k) = [0 \ 1 \ 1][1 \ 2 \ 4]^t = 6$

Since the value is positive an update is not carried out. So,

$\theta^t(k+1) = \theta^t(k) = [0 \ 1 \ 1]$

For iteration 2,

$\theta^t X(k) = [0 \ 1 \ 1][1 \ 3 \ 3]^t = 6$

Since the value is positive an update is not carried out. So,

$\theta^t(k+1) = \theta^t(k) = [0 \ 1 \ 1]$

For iteration 3,

$\theta^t X(k) = [0 \ 1 \ 1][-1 \ -6 \ -12]^t = -18$

Since the value is negative an update is carried out as follows,

$\theta^t(k+1) = \theta^t(k) + \alpha X(k) = [0 \ 1 \ 1] + 1 * [1 \ 6 \ 12]^t = [-1 \ -5 \ -11]$

For iteration 4,

$\theta^t X(k) = [-1 \ -5 \ -11][-1 \ -8 \ -10]^t = 151$

Since the value is positive an update is not carried out.

c. Illustrate 4 iterations of relaxation with $b = 1$

At first all the samples are augmented by 1 and samples for class 2 are multiplied by -1. So

For iteration 1,

$\theta^t X(k) = [0 \ 1 \ 1][1 \ 2 \ 4]^t = 6$

Since the value is greater than b(=1) an update is not carried out.

For iteration 2,

$\theta^t X(k) = [0 \ 1 \ 1][1 \ 3 \ 3]^t = 6$

Since the value is greater than b(=1) an update is not carried out.

For iteration 3,

$\theta^t X(k) = [0 \ 1 \ 1][-1 \ -6 \ -12]^t = -18$

Since the value is less than b(=1) is θ updates as follows,

$$\theta(k+1) = \theta(k) - \alpha\left[\frac{\theta^t(k)X(k)-b}{|X(k)|^2}\right]X(k) = [0\ 1\ 1] - 0.1 * (-9.5) * [-1\ -6\ -12]$$

$$= [-0.95\ -5.7\ -11.4]$$

For iteration 4,

$$\theta^t X(k) = [-0.95\ -5.7\ -11.4]\,[-1\ -8\ -10]^t = 160.55$$

Since the value is greater than b(=1) an update is not carried out.

d. Illustrate 4 iterations of Ho-Kashyap algorithm with $b = [1\ 1\ 1\ 1]^t$

As before, all the samples are augmented by 1 and the samples of class 2 are multiplied by -1.

Let's evaluate the value of $X^{\#} = \left[X^t X\right]^{-1} X^t$

0.799	0.699	-0.153	-0.353
-0.168	0.082	-0.212	0.288
0.032	-0.118	0.188	-0.122

The initial value of θ is $X^{\#}b = [2.004\ -0.162\ -0.162\ ]^t$

For iteration 1,

$$e(1) = [0.032\ 0.032\ -0.087\ -0.087]^t$$

$$b(1) = [1.064\ 1.064\ 1.000\ 1.000]^t$$

$$\theta(2) = [2.100\ -0.168\ -0.168]^t$$

For iteration 2,

$$e(2) = [0.028\ 0.028\ -0.076\ -0.076\ ]^t$$

$$b(2) = [1.070\ 1.070\ 1.000\ 1.000]^t$$

$$\theta(3) = [2.109\ -0.168\ -0.168]^t$$

For iteration 3,

$$e(3) = [0.031\ 0.031\ -0.085\ -0.085]^t$$

$$b(3) = [1.076\ 1.076\ 1.000\ 1.000]^t$$

$$\theta(4) = [2.118\ -0.169\ -0.169]^t$$

For iteration 4,

$$e(4) = [0.028\ 0.028\ -0.076\ -0.076]^t$$

$$b(4) = [1.081\ 1.081\ 1.000\ 1.000\ ]^t$$

$$\theta(5) = [2.125\ -0.169\ -0.169\ ]^t$$