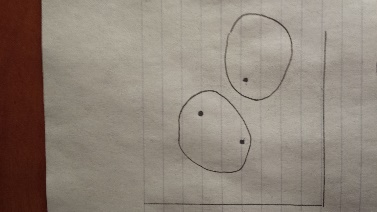
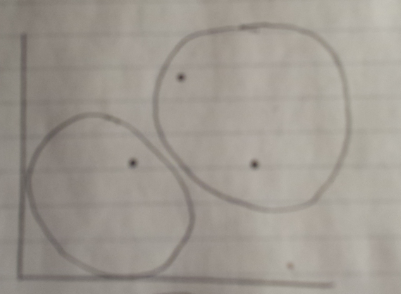
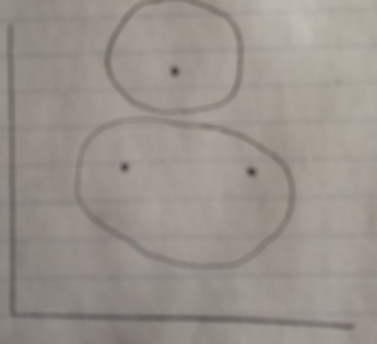
Nathan Moeller

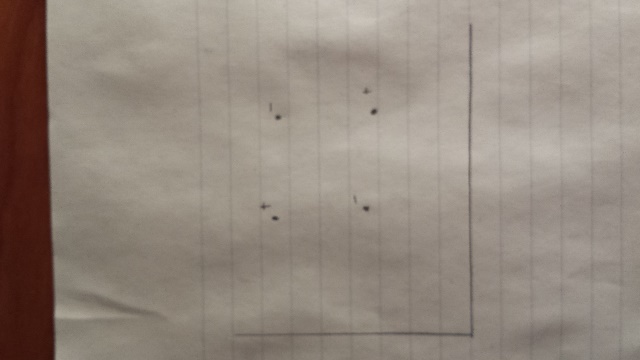
CSci 5521 Machine Learning

Homework 1

1. The definition of VC dimension is the upper bound on the number of data points the classification function can shatter. Our classification function is: sin(abs(x – c) –r). This function results in a circle. So, to find the VC dimension of sin (abs(x – c ) – r), the VC dimension of a circle must be derived. It is obvious that the VC dimension of a circle is at least 3 because of the following diagrams.

No matter what each point is labeled, the circle classifier can shatter all of the points. However, this is not true when another point is added. This can be seen in the diagram below.



As shown in the above diagram, the diagonal points are the same class. In this configuration, it is not possible to correctly classify all four points with a circle. Because of this, the VC dimension of sin(abs(x – c) – r) is 3.

2. P(x | C) = p^x \* (1-p)^(1-x). This is the Bernoulli density function. This would be substituted in the classification rules below.

Based off Bayes theorem, we can calculate the posteriors. The evidence is omitted because it is not needed.

P(C1 | x = 1) = P(x = 1 | C1) \* P(C1)

P(C2 | x = 1) = P(x = 1 | C2) \* P(C2)

P(C1 | x = 0) = P(x = 0 | C1) \* P(C1)

P(C2 | x = 0) = P(x = 0 | C2) \* P(C2)

So when x = 1,

P(C1 | x = 1) > P(C2 | x = 1) classify as C1

P(C2 | x = 1) > P(C1 | x = 1) classify as C2

When x = 0,

P(C1 | x = 0) > P(C2 | x = 0) classify as C1

P(C2 | x = 0) > P(C1 | x = 0) classify as C2

1. Pij = P(xj = 1 | Ci ) for i = 1,2 and j = 1,2,3,4,….D

The above specifies D dimensional independent Bernoulli density functions. To classify a sample into class 1 or 2, we must solve for:

P(Ci | x1, x2, x3….xD) = ?

Because each variable is independent, we can multiply each one together

P(C1 | x1, x2, x3….xD) = P(C1 | x1) \* P(C1 | x2) \* P(C1 | x3) \* ….. \* P(C1 | xD)

P(C2 | x1, x2, x3….xD) = P(C2 | x1) \* P(C2 | x2) \* P(C2 | x3) \* ….. \* P(C2 | xD)

Using the density function for an individual variable, the above multiplications can be computed. Next, all we have to do if multiply by the prior and compare which is larger. Again, the evidence is not included because it is not needed.

If,

P(C1 | x1, x2, …..xD) \* P(C1) > P(C2 | x1, x2, ….xD) \* P(C2) classify as C1

P(C2 | x1, x2, …..xD) \* P(C2) > P(C1 | x1, x2, ….xD) \* P(C1) classify as C2

**For P(C1) = .1, P(C2) = .9**

Sample {1, 1} = Class2

P(Ci | x1, x2) = P(x1 = 1 | Ci) \* P(x2 = 1 | Ci) \* P(Ci)

P(C1 | x1 = 1, x2 = 1) = (.7)(.2)(.1) = .014

P(C2 | x1 = 1, x2 = 1) = (.4)(.8)(.9) = .288

Sample {1,0} = Class2

P(C1 | x1 = 1, x2 = 0) = (.7)(.8)(.1) = .056

P(C2 | x1 = 1, x2 = 0) = (.4)(.2)(.9) = .072

Sample {0, 1} = Class2

P(C1 | x1 = 0, x2 = 1) = (.3)(.2)(.1) = .006

P(C2 | x1 = 0, x2 = 1) = (.6)(.8)(.9) = .432

Sample {0, 0} = Class2

P(C1 | x1 = 0, x2 = 0) = (.3)(.8)(.1) = .024

P(C2 | x1 = 0, x2 = 0) = (.6)(.2)(.9) = .108

**For P(C1) = .5, P(C2) = .5**

Sample {1, 1} = Class2

P(C1 | x1 = 1, x2 = 1) = (.7)(.2)(.5) = .07

P(C2 | x1 = 1, x2 = 1) = (.4)(.8)(.5) = .16

Sample {1,0} = Class1

P(C1 | x1 = 1, x2 = 0) = (.7)(.8)(.5) = .28

P(C2 | x1 = 1, x2 = 0) = (.4)(.2)(.5) = .04

Sample {0, 1} = Class2

P(C1 | x1 = 0, x2 = 1) = (.3)(.2)(.5) = .03

P(C2 | x1 = 0, x2 = 1) = (.6)(.8)(.5) = .24

Sample {0, 0} = Class1

P(C1 | x1 = 0, x2 = 0) = (.3)(.8)(.5) = .12

P(C2 | x1 = 0, x2 = 0) = (.6)(.2)(.5) = .06

**For P(C1) = .9, P(C2) = .1**

Sample {1, 1} = Class1

P(C1 | x1 = 1, x2 = 1) = (.7)(.2)(.9) = .126

P(C2 | x1 = 1, x2 = 1) = (.4)(.8)(.1) = .032

Sample {1,0} = Class1

P(C1 | x1 = 1, x2 = 0) = (.7)(.8)(.9) = .504

P(C2 | x1 = 1, x2 = 0) = (.4)(.2)(.1) = .008

Sample {0, 1} = Class1

P(C1 | x1 = 0, x2 = 1) = (.3)(.2)(.9) = .054

P(C2 | x1 = 0, x2 = 1) = (.6)(.8)(.1) = .048

Sample {0, 0} = Class1

P(C1 | x1 = 0, x2 = 0) = (.3)(.8)(.9) = .216

P(C2 | x1 = 0, x2 = 0) = (.6)(.2)(.1) = .012

1. Table of error rates for the validation set:

|  |  |
| --- | --- |
| **Prior** | **Error Rate** |
| P(C1) = .1 P(C2) = .9 | 0.16854 |
| P(C1) = .2 P(C2) = .8 | 0.16854 |
| P(C1) = .3 P(C2) = .7 | 0.16854 |
| P(C1) = .4 P(C2) = .6 | 0.16854 |
| P(C1) = .5 P(C2) = .5 | 0.16854 |
| P(C1) = .6 P(C2) = .4 | 0.16854 |
| P(C1) = .7 P(C2) = .3 | 0.20225 |
| P(C1) = .8 P(C2) = .2 | 0.20225 |
| P(C1) = .9 P(C2) = .1 | 0.20225 |

Error rate for test set: 0.1573