Some algebra done with mathematica CAS

# Problem 1

We have:

if we want bound:

we need

So we need at least points which is closest to 1000 so the answer is B

# Problem 2

We use the same equation as (1) but with

So we need at least 1301 points which is closest to 1500 so the answer is C

# Problem 3

We use the same equation as (1) but with

So we need at least 1761 points which is closest to 2000 so the answer is D

# Problem 4

We know that is the most data points that a hypothesis set can shatter regardless of their location in . Thus, the smallest break point for a hypothesis set is . We know that for a 1D perceptron we would have , and for 2D perceptron we would have . Thus we conjecture that a 3D perceptron would have .

We will prove that below:

We must prove that and that

For , consider points in shattered by perceptron formed around:

and any:

We look to find s.t.

Which we can do by

Thus, we can shatter these points

For the direction we must show that we can’t shatter any set of points

So we take any 5 points:

Since we have more points than dimensions we know at least 1 point is a non-trivial linear combination of the other 4 points or:

and

Consider the dichotomy where ’s with in the above equation have , and has .

Since:

we know:

So since , then which means that:

So since has a forced value, we know we can’t achieve all dichotomies.

Thus .

We have and , thus and since we have a min which is B.

# Problem 5

We will use what we know from p. 49 of the book:

If for some value , then:

for all

which is a polynomial in with degree . Thus, any possible growth function must be bondable by polynomial in .

(iv)

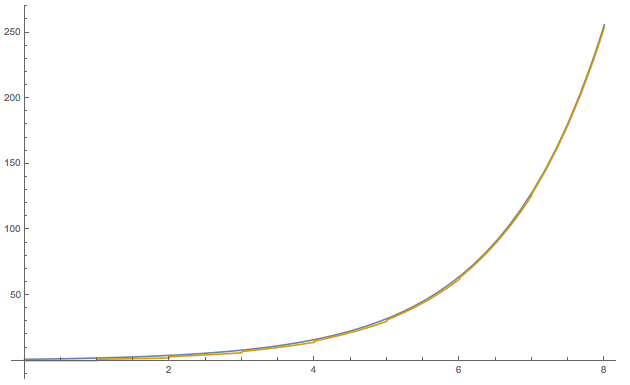
By this we know that is not a possible growth function because if there is some breakpoint than the growth function must be polynomial in and if there is no breakpoint than the growth function is so it can’t be . Since we know that it must be bounded by a polynomial to be a valid growth function but it is not.

(i)

We know this is a valid growth function since it is exactly the growth function of the positive rays model mentioned on p. 43.

(iii)

We graph in blue next to in orange



These graphs tell us a few things. Firstly, if where a growth function of a hypothesis set it would have a break point at which doesn’t really make sense for starters because there is only ONE possible way for 0 points to be classified therefore it would seem impossible that a hypothesis set could have a breakpoint at . Further, inspection tells us that althoughseems to remain less than , it is clearly growing alongside which would indicate that it is growing exponentially which would be in violation of the above condition.

Further, we can evaluate for some where and compare it to . We choose which by the graph clearly satisfies . So we have:

and

/Users/eli/Desktop/Screen Shot 2017-10-09 at 6.54.26 PM.png

As we can see, the possible quickly overwhelms the bound meaning that it is an impossible growth function.

Additionally, we know that the expression is in fact exponential with respect to since order of the polynomial is . Which makes it exponential in and thus not a valid growth function since we also know it is not equal to .

(v)

We know that this is a valid growth function as it is the max growth function corresponding to a hypothesis set that can produce all possible dichotomies. This is the growth function of any hypothesis set with breakpoint .

(ii)

we examine:

We know this is a possible growth function as it is exactly the growth function as the positive intervals model mentioned on p44 of the book.

So we have (i), (ii), (v) are valid which means the answer is B

# **I’m doing 7 before 6 because the answer to 7 helps to find the answer of 6**

# Problem 7

We can determine the growth function by thinking about the ways in which we can divide two separate intervals.

We can separate into three cases:

In the first case, the intervals overlap to some extent which means together, they are effectively one interval. As such we can choose a start and an end point to this combined interval. With data points we have spots to put an endpoint to the interval. Thus, we have .

In the second case, the intervals as perfectly disjoint. In this case we choose 4 endpoints to define 2 intervals. We can say that there are , but we must be clear to note that for any selection of 4 endpoints, you may only connect the leftmost with the second leftmost and the rightmost with the second rightmost. This is because if you did otherwise you would have an overlap which was covered in the first case. The important takeaway is that for any 4 endpoints you can assign exactly 1 configuration of disjoint intervals.

The final case is the trivial case when all four endpoints are chosen in the same place which makes . This adds one dichotomy.

Thus, the overall answer is which is C

# Problem 6

Given our answer for 7 we can look for the value of N where

We have

for

and

for

Thus

and the smallest breakpoint is so we have C

# Problem 8

For starters, we know from (6) that the breakpoint for is so the answer must be D

We will also look at the case for confirmation.

If we have 3 intervals, we have three cases,

First case, we have 3 overlapping intervals in which case we are dealing with the positive intervals model so we have dichotomies.

Second case, we have 2 overlapping intervals and a 3rd disjoint interval so we are essentially dealing with the second case in the 2 intervals model which gives us more dichotomies

Third case, we have 3 disjoint intervals in which case we must choose 6 endpoints and we have

So for we have:

Its not hard to see that the general case is:

back to the case we have

for

but for

So is our breakpoint which is in agreement with for .

Again, the answer is D

# Problem 9

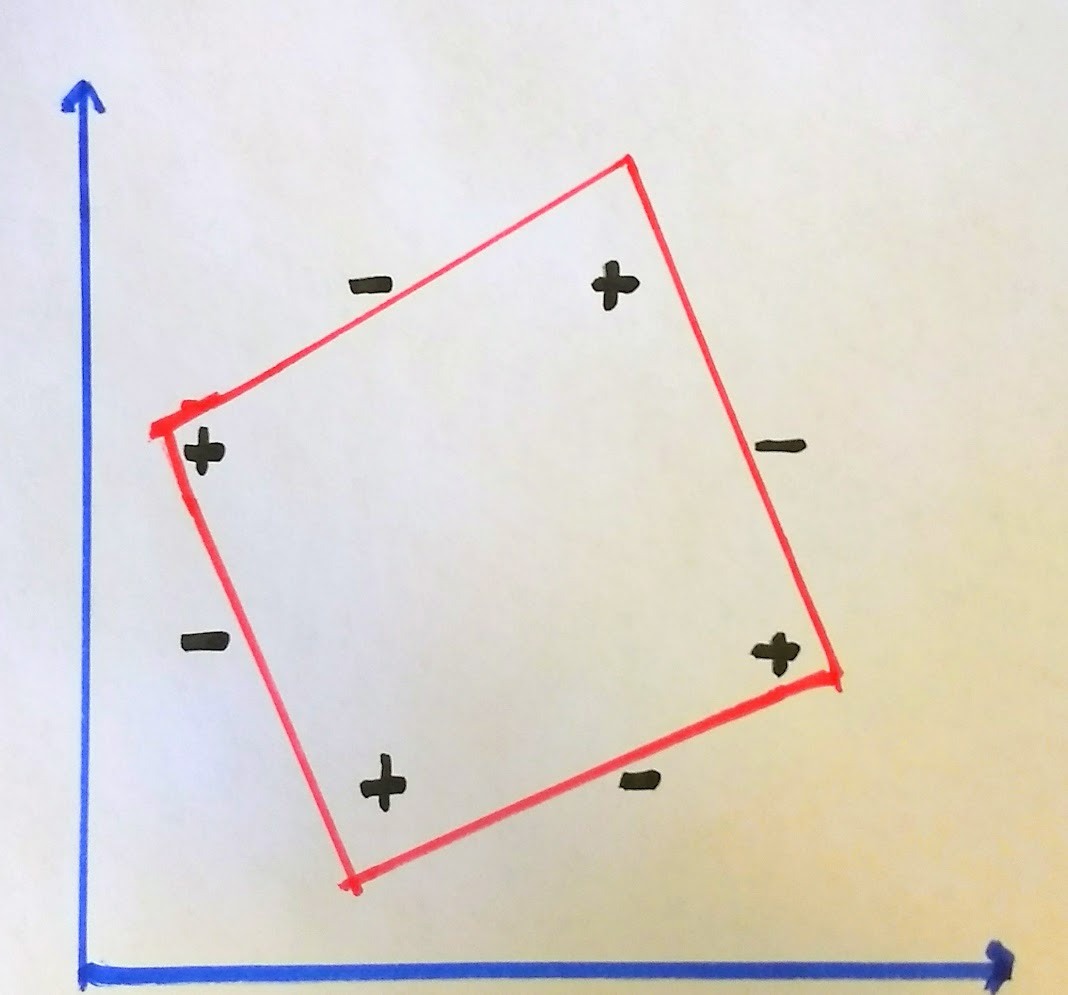
We know that in order to shatter a set of points you must be able to pick any arbitrary subset of points and enclose them in the triangle. We will be focusing on the same example arrangement as used in lecture with convex sets as we know them to be the “easiest to dichotomize” with convex regions. Thus, any failure to do so in that case will extend to the general case.

We know that we can select an arbitrary set of points on a convex set for any because we can enclose a single point in a small triangle, any 2 points in a flattened triangle, and any 3 points in a triangle with vertices near the points.

We know that we can be sure to exclude no more than 3 specific points (one with each edge), we can often exclude more than 1 per edge however we can only **guarantee** that we can exclude 1 per edge. This is possible when no three points are collinear which is the case in the convex set formation.

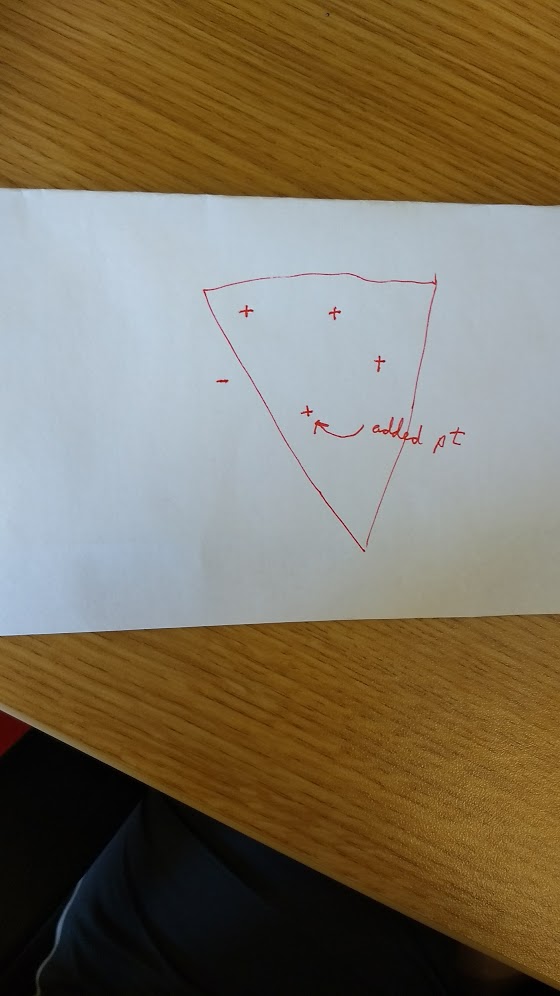
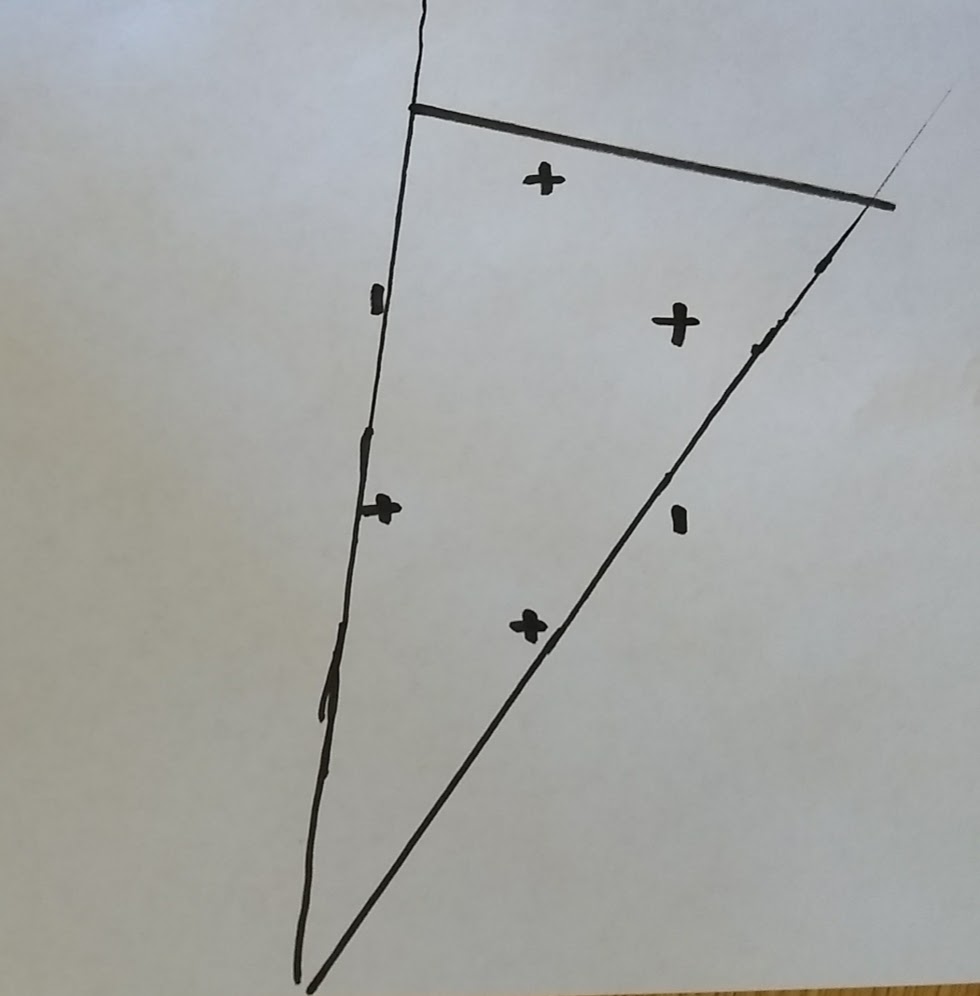
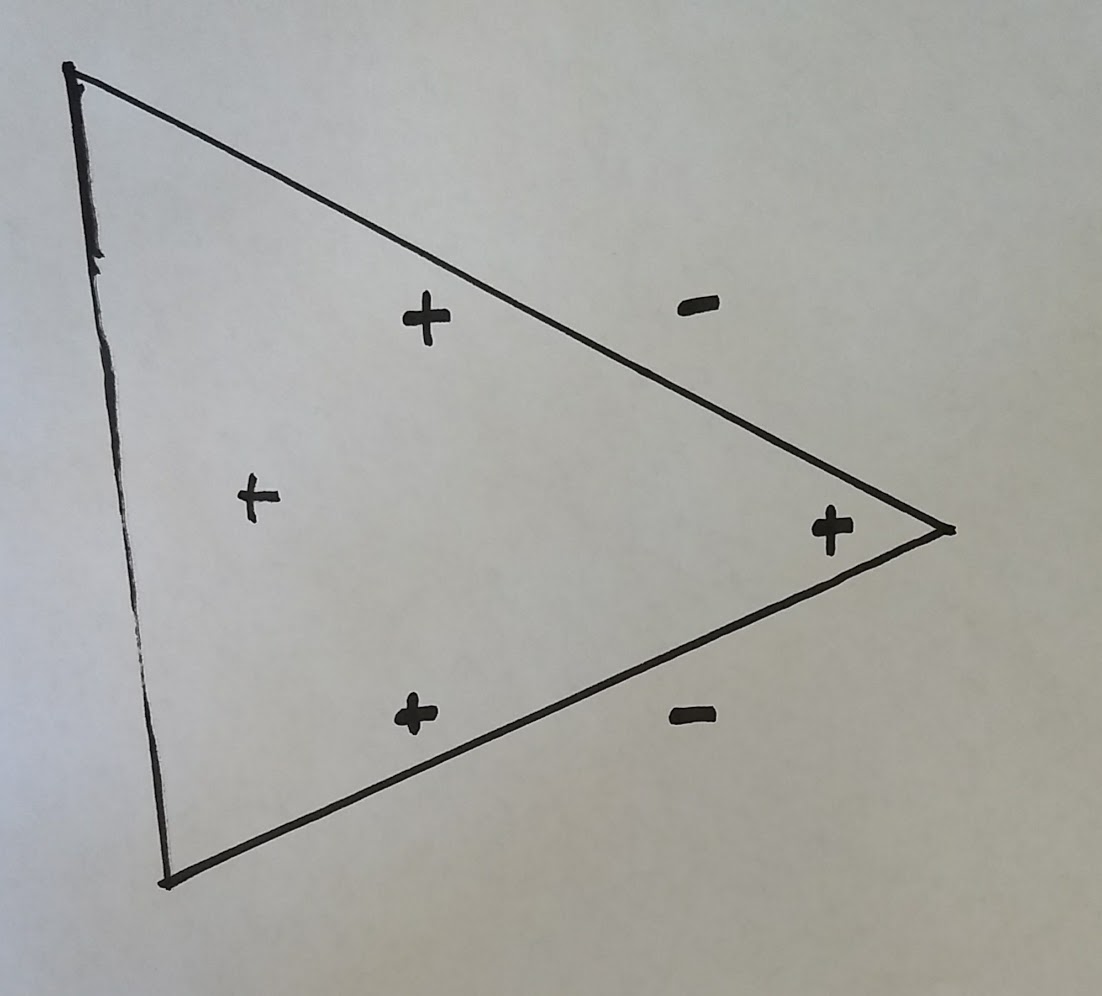
This means that for a set of points we know that we can select any points in a triangle by excluding 3,2,1 or 0 of the points in the data set by placing an edge of the triangle just inside of the point.

For , we know from above that we can choose any points arbitrarily with the triangle thus we can shatter .

We find our first break point at as shown by the following impossible dichotomy

Thus the largest that can be shattered by the triangles hypothesis is 7 which is D

Additional examples are given for clarity:



# Problem 10

Since the classifying feature of a point is just whether or not the radius: is within some interval , we can reduce the concentric circles model to the “positive intervals” model that we discussed in lecture. We do this without loss of data relevant to classification.

Instead of the 2D feature we take the 1D feature

Thus for such points we have partition points to put two endpoints to the interval that would result in distinct dichotomies as in the positive. So from that choice we would have . However, unlike the positive intervals model, we have the possibility that the two endpoints can be in the same partition point in which case all have so that adds one more dichotomy.

So the answer is B