ML from Data – HW 5

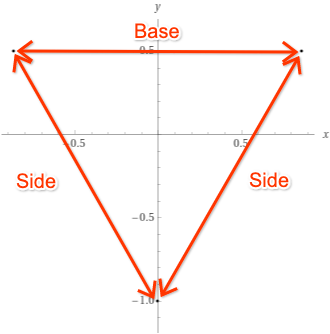
319238119, 207576463

1. Full rational variety
2. We saved 10 multiplication operations by using instead
3. ,

From 1: From 2:

From 1+2+3:

1. Given [this piazza post](https://piazza.com/class/l09q5swe2yf37h?cid=213), In this solution we assume the triangle is of the following form, even though this triangle is not upright as stated in the question. This detail is not critical for the solution complexity but changes some of the details in it.



Algorithm to calculate an origin-centered equilateral triangle given a dataset of points of size in and as defined in the question:

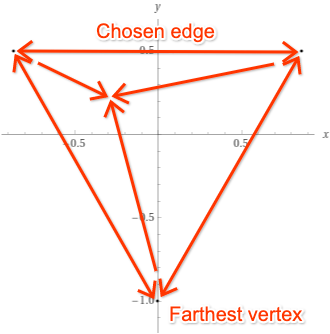
1. For every point in :
   1. Calculate the origin-centered equilateral triangle which contains on its perimeter.
   2. Return if all the other points are contained in it.

We returned the smallest triangle that contains all the points and hence, given that the data is in fact correct, .

There are two points left to explain:

1. How to implement step 3.1)a) – Given a point , how to calculate an origin-centered equilateral triangle that contains on its perimeter?

We start by finding which edge of the triangle should contain the point. We do this by finding the farthest triangle vertex from the point, which indicates that the point should be on the edge that is in front of this vertex, as demonstrated in the following diagram.



At this point we know which edge contains the point . We also can calculate easily the slopes values of all the triangle edges, because the triangle is origin-centered and equilateral. Using simple math skills we can get the all the 3 linear lines of the form that form the triangle in a fixed amount of steps denoted as (we don’t write the full logic because it’s long, has edge cases and is not the primary purpose of this assignment).

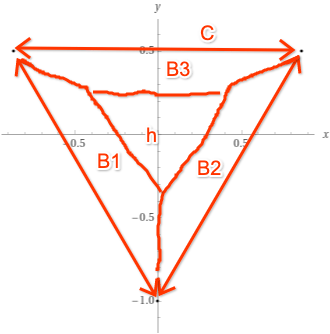
1. How to implement step 3.1)b) - Given a triangle defined by 3 linear lines of the form of which one is the triangle base and the others are its sides and a point , how do we check whether the triangle contains ?

We simply check that is **above** the triangle sides and **below** the triangle base. We do that by putting the value of the point in the line formula and comparing the values of . This is a simple check that requires a small fixed amount of actions denoted by .

Building a triangle for a point in the set costs and checking whether all the other points are within it costs . We do these actions for all the points and therefore the time complexity of this algorithm is which is polynomial on the size of the input data set .

We now would like to calculate a bound on the number of samples that ensures with probably that . The bound will be of the form .

Suppose is a probability function defined on the surface.

We define the sets to be inner trapezoids, each shares an edge with the triangle and their upper bases form a smaller inner triangle s.t. .

If a dataset visits each , then we know that . We want to block the probability that does not visit all s.

We know that:

Therefore: .

Therefore:

From that we get:

Therefore, in order to know with probably that , the training set size has to be at least .

1. The data we have is: , hence:

with confidence.

The true error that can be expected is up to 22.479%

1. 