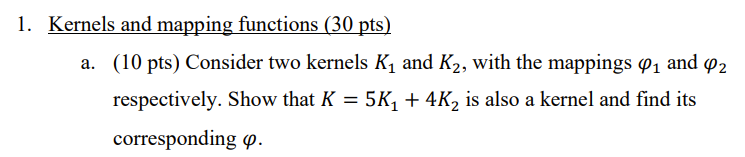
HW5:

1a.

In terms of the mapping function 𝜑, the relationship of a kernel to its mapping is given by 𝐾(x, y) = ⟨𝜑(x), 𝜑(y)⟩. For 𝐾1 and 𝐾2, we can write this as 𝐾1(x, y) = ⟨𝜑1(x), 𝜑1(y)⟩ and 𝐾2(x, y) = ⟨𝜑2(x), 𝜑2(y)⟩.

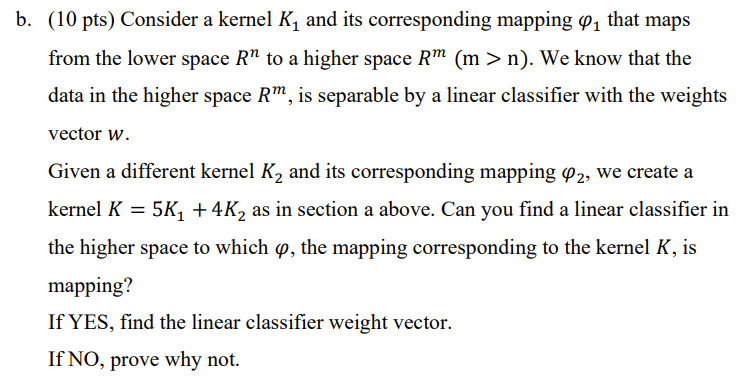
Define 𝜑(x) = [√5𝜑1(x), √4𝜑2(x)]

Thus, we can express 𝐾 = 5𝐾1 + 4𝐾2 as follows:

𝐾(x, y) = 5𝐾1(x, y) + 4𝐾2(x, y)

= 5⟨𝜑1(x), 𝜑1(y)⟩ + 4⟨𝜑2(x), 𝜑2(y)⟩

Therefore, the resulting kernel 𝐾 corresponds to the mapping 𝜑(x) as described above.



1b. it is known that the data in the higher space is separable by a linear classifier with weights vector w.

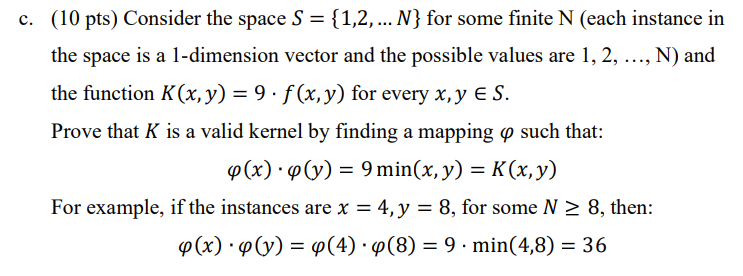
Hence we classify in the next form:

Now we’ll find linear separator for .

maps to dimension X.

We will choose .  
We will define vector W for our separator W = ()

Hence, from the dot product and as seen previously if it’s > 0 , we classify 1, else -1.



1c.

Consider the following mapping function φ(x):

For a given x, φ(x) is a N-dimensional vector where the first min(x) entries are sqrt(9) and the remaining entries are zero.

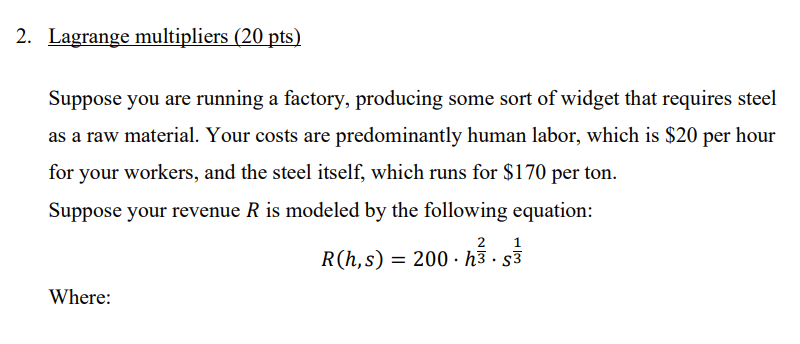
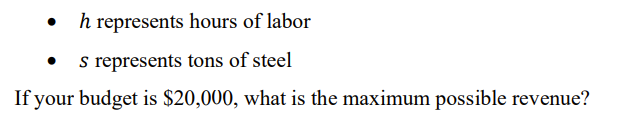
Now, let's compute the dot product φ(x)·φ(y) and see if it equals K(x, y) for all x, y in S:

The dot product is computed by multiplying corresponding entries in each vector and then summing those products. Because the only non-zero entries in the vectors φ(x) and φ(y) are the first min(x) and min(y) entries respectively, the dot product essentially becomes the sum of the products of these non-zero entries.

If x <= y, then the non-zero entries in φ(x)·φ(y) are the first x entries. For those terms, φ(x)[i] = φ(y)[i] = sqrt(9). The dot product is then x \* 9 = 9 \* min(x, y).

If y < x, then the non-zero entries in φ(x)·φ(y) are the first y entries. For those terms, φ(x)[i] = φ(y)[i] = sqrt(9). The dot product is then y \* 9 = 9 \* min(x, y).

In both cases, φ(x)·φ(y) = 9 \* min(x, y) = K(x, y), confirming that K is a valid kernel with the proposed feature mapping φ.



2. The equation for revenue R is given by R(h, s) = 200 \* h^(2/3) \* s^(1/3), and we have a budget constraint of 20 \* h + 170 \* s = 20000, where h represents hours of labor and s represents tons of steel.

We can use the method of Lagrange multipliers to find the maximum revenue. In this case, the Lagrange function L(h, s, λ) is given by:

L(h, s, λ) = 200 \* h^(2/3) \* s^(1/3) + λ \* (20h + 170s - 20000)

Taking the partial derivatives of L with respect to h, s, and λ and setting them equal to 0, we get the following equations:

∂L/∂h = (200 \* 2/3 \* h^(-1/3) \* s^(1/3)) + 20λ = 0 (equation 1)

∂L/∂s = (200 \* 1/3 \* h^(2/3) \* s^(-2/3)) + 170λ = 0 (equation 2)

∂L/∂λ = 20h + 170s - 20000 = 0 (equation 3)

We can solve this system of equations to find the optimal values of h and s.

S = 2000/51 ≈ 39.21 tons

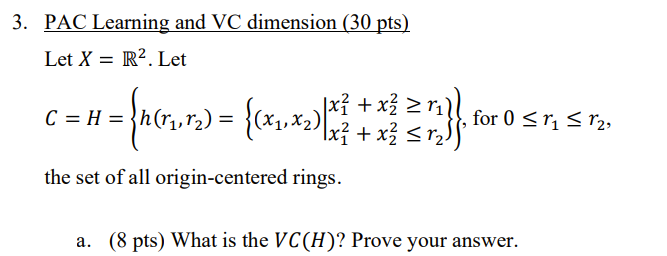
h ≈ 2000/3 ≈ 666.66 hours

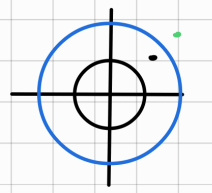
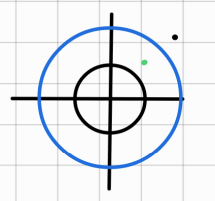
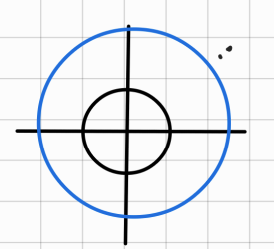
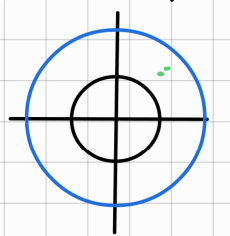
So, the maximum revenue occurs with about 666.66 hours of labour and 39.21 tons of steel.

To find the maximum revenue, we substitute these values into the revenue function:

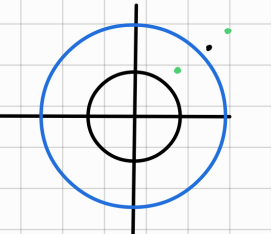
R(666.66, 39.21) = 200 \* (666.66)^(2/3) \* (39.21)^(1/3) ≈ $51,852.0

So, the maximum possible revenue is about $51,852.0

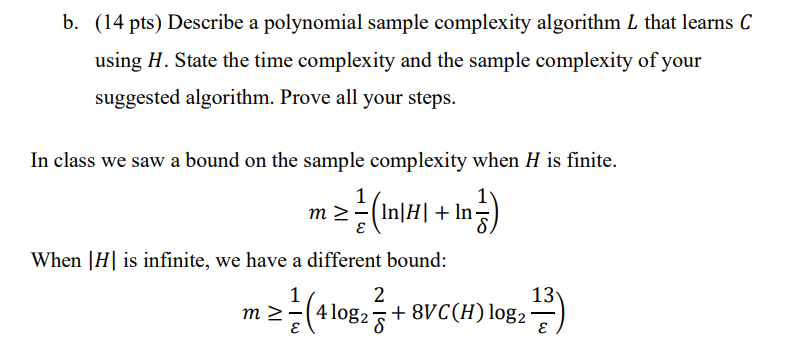


3a. VC(H)=2. First we will see that   
There is a separation for every dichotomy:  


Now we will prove that , let set of 3 different points .

If the 3 points are colinear, we will choose labeling (+ - +)   
And that is impossible.

If the three points are convex hull, we choose the labeling + - + again and that is impossible to find a linear separator for them. 



3b.

Algorithm:

We will go through training data D.

We will find the point that is classified as positive (1) and that its distance from the origin center is minimal, i.e.,

and another point that is classified as positive (1) and that its distance from the origin center is maximal, i.e.,

We will draw the rings according to , hence return h=L(D) such that

* Note: Different training datasets will cause different results.

The algorithm is polynomial because it will cost O(m) to find the max and min out of m samples.

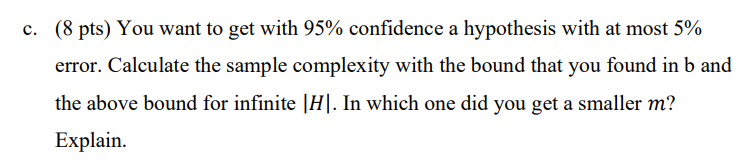
Correctness of the algorithm – if we block all positive points with the minimal length up to the maximal length, then the classification was correct.

Now we will find the sample complexity:

Define – the probability that one point is located inside circle.

– the probability that one point is located outside circle.

Hence the sample size



3c.

Define – the probability that one point is located inside circle.

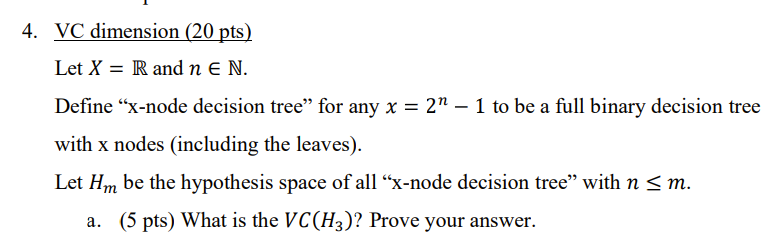
– the probability that one point is located outside circle.

Hence the sample size

Substitute , in the two bound formulas

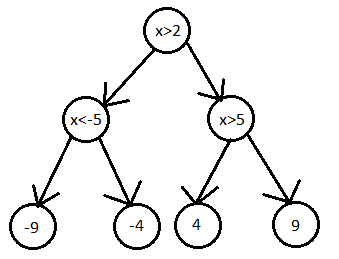
1. instance 2992
2. \* m(0.05,0.05)=

We got two bounds for number of samples. Choose the bound 147 since it is tighter.



4a. VC(H)=4

Firstly, for 4 points: -9, -4, 4, 9 we build the decision tree:



We can assign 2 options for every leaf which gives us different dichotomies.

We show that VC(H)<5.

Assume towards contradiction that we can shatter 5 points, denote by .

But our tree has 4 leaves and from the pigeonhole principle there exists a leaf that classifies more than one point, and as we need to show a dichotomy for every label assigning, there exists a case where the classified points contradict.

Hence, while that are classified by the same leaf and then we can classify + to and – to . We didn’t reach a valid dichotomy, contradiction.



4b. we will show that VC ()

First, we will show that

We look at points . There exists a tree that classifies each point to a different leaf. And overall, there are leaves.

We can assign different dichotomies.

We will show that

Assume toward contradiction that we can shatter the points

We will use pigeonhole principle, hence there are at least two points which will be at the same leaf, and we could label them as different classes, and therefore we get a contradiction.