**Part 2 Math & Statistics**

**a) You have 10,000+ variously sized polyhedrons, all plotted individually in a cartesian space. Discuss a possible methodology that could be used to remove duplicate polyhedrons and be reasonably assured you have a final set of unique structures.**

A polyhedron is a 3-d solid made up of polygon faces.  Polyhedron have vertices, edges and faces.  Two vertices create an edge, multiple edges connected make a face.

Consider we have a list of N polyhedrons.  (If these were randomly located, we would need to traverse the entire cartesian space which is infinite, so decided to start with the idea that polyhedron location(s) are known).

To compare duplicate polyhedrons, we will need to traverse the complete list of polyhedrons.  This creates close to a factorial computational time.  (Depending on the number of actual duplicates there are).  We will take polyhedron1 & compare polyhedrons 2 through N. If polyehdron2 is not a duplicate, we then compare it to polyhedron 3 through N(remaining).

To save time we will rule out polyhedrons that do not have the same number of vertices.

(If we know the number of edges & number of faces as if we had a list of polyhedron objects then we could compare these as well.)

If the number of vertices in two polyhedrons match, we will check to see if there exists one vertex edge set in polyhedron2 that matches the first vertex edge set in polyhedron1.  (An edge set will be considered the group of edges that share a single vertex.  If you think of corner Z in a cube, the three edges connected to corner Z will be an edge set.)  Two equal edge sets should have the same number of connected edges of equivalent lengths.  If the first edge set in polyhedron1 does not have a match in polyhedron2, move on; it's not a duplicate.  If there is an edge set that matches, now traverse through each edge set of polyhedron1, making sure polyhedron2 has a match. These should be recorded as corresponding edges.  If any edges in polyhedron2 does not have a corresponding edge, then not a duplicate.  (Could have a case where polyhedron1 has a match completely, but polyhedron2 does not.) If we are left with only the corresponding edges, we move on to the next part.

Now that we have a list of corresponding edges, we must check their angles.  By using dot product between two edges / edge1 magnitude \* edge2 magnitude we can find each angle between edges in an edge set.  If we have an edge set (e1,e2,e3) we must find the angles between e1e2, e1e3,e2e3.  Then we compare these angles to the corresponding edge set.  Each angle must have an equivalent angle while taking corresponding edge length into account.  Want to make sure the edge lengths that matched up, their corresponding angle is the one matching up.

If these checks are all passed then the structures on these edges need to be compared.  Depending on the information provided we either need to construct the face by traversing adjacent edges, otherwise each edge will need knowledge of the faces it belongs to.  We must compare the corresponding edges, corresponding faces.

If each check is passed, we remove polyhedron2 aka the duplicate from the list to leave only unique polyhedrons.

This process should remove any duplicate polyhedrons if duplicate polyhedrons consider where vertices & edges are placed.  It is possible to have say two cubes to be constructed that are the same in outside structure, but built in different ways.  Cube1 has all square faces, cube2 has 2 rectangles connected forming a square for each face.  The vertex, edge pattern of cube2 will differ from cube 1.  If polyhedrons are duplicates based only on overall structure, then faces in the same plane that are connected must be taken into consideration to then compare with other structures.  (I will leave it at case1 for now)

**b) You have an n-dimensional surface that represents the solution space of your problem. A dense sampling of this space is not computationally feasible.**

**How can the global minimum of this surface be determined?**

**How can you be sure you are not in a metastable state?**

Generally one could use stochastic gradient descent to find the global minimum of the surface, but due to computational restraints we can use batch gradient descent where we choose an effective batch size, one that is below computational thresholds.  Basically, you want to keep the randomness as high as possible to prevent getting stuck in metastable states are.  Other ways to avoid metastable states are to randomize initialization weights, have a dynamic learning rate, or apply a momentum factor.

How can you be sure you are not in a metastable state?

Seems similar to the question `what came first, the chicken or the egg? ` If you know the global minimum, you could be sure you are not in a metastable state.  Best way to know is through randomization and comparing solutions.  The more computations & local minima/saddle points you can compare, the closer you are to having the global minimum.

**PART 3 BUSINESS INSIGHT**

Question 3: Business Insight

You are hired as a data scientist by a general contractor that constructs townhome communities all over Pennsylvania. These communities are of various sizes as well as various price and quality.

**a) What are some considerations to be made when determining a location for the construction of a new townhome community?**

Land Specific:

What is the actual cost of the land?

How much land do you need to incorporate everything that comes with a “community” (Walking Trails, Pool, Park, Tennis Court, Occupant Parking, Visitor Parking, Playground, Communal Centers, etc.)

The type of land and how that affects the material needed to build quality structures?

Will the house be prone to natural disasters? Flooding, Hail, Hurricanes etc.  How does it affect necessary materials?

Can building permits be obtained?  What are the property taxes on the land?

Business/Attraction

What businesses & destinations are close by?  Are there things to do that are close by?  What type of person/household do these destinations cater to?  (Family Oriented, Nightlife, Adventurers, etc.)

What jobs and income levels are accessible to the location?  Overview of who are the possible buyers, how much they can afford, do they have stable income, what are the commute times to various companies & industries, what are the job opportunities close by?  Can the potential buyers secure credit to purchase a townhome?

Are these homes going to be primary or secondary residences?

Are these homes going to be sold, or rental properties?  Does there need to be a property management team?

Building Specs:

Based on necessary materials needed to build, and potential buyer income prospects, how do we cater design & price point to them?  Make sure we are not outpricing our key buyer demographic.

What level of variation is needed between townhomes (design, curb appeal, interior design, etc) are possible/needed?  What amenities are included with the home? What customizations can be done?  What are build times between different design plans?  What is expected human capital cost to complete the job?  What delays can be expected in this region?  How many actual townhomes are being built?

Home Sales:

How do we market the community?  How do we sell the houses?  Are there realty groups we have partnerships with nearby?  What level of branding does our company have? How quickly do various buyer demographics close on sales? How long can we expect the sales cycle to last for?  What are the costs associated with a sales & marketing team?  What is overall human capital expenditure?

Population/demographics: How many people are in the region or close to it?  Is the population increasing or decreasing?  What are various income levels?  What are job opportunities?  What is the stability of these jobs?  What credit do these individuals possess?  What price point can credit/income reach?  How many people are there vs homes in the area?  How many people live in a single home? How many generations live in a single?

Townhome Acceptance:

How many Townhome communities are nearby?  What quality are they?  How many are filled/vacant?  Owned vs rental?

Are buyers looking for townhomes or detached homes?  What is the breakdown between the two near the prospective location(s)?

b) Incorporating these considerations, how would you approach designing a model that would output the current viability of constructing a townhome community at a location?

How would you approach designing a forecast for the future demand of new townhome communities in Pennsylvania?

**Viability Model:**

Would look to acquire relevant data pertaining to the questions I have above.  Depending on data I could gather access to would look to create a model to optimize profit based on how costly it is to develop, build & sell at the location vs expected revenue the properties can generate.  First thought is to do this at the property level or townhome level coupled with buyer demographics.  Each townhome has an associated cost & value with it.  This allows for a larger dataset than just prior communities built, and gives insight on how large the community should be.  The townhome should have an extensive feature set of various cost centers, the sales price point, the associated buyer demographics, the design of townhomes & community.  Ideally would want to see how various buyer demographics interact with various townhomes/communities in the past, to predict how changes to prospective communities can increase profitability.  (I.e. if high earning `households` will not purchase from previous communities why not? Then are there enough of these `households` to build a higher quality community, or should a different demographic be built for) So the model needs to have some combination of buyers’ vs build types, or maybe a model for each.  Again, dependent on data & possible data structure.

**Future Demand:**

To forecast future demand, I would base the model on prior townhome community results, as well other housing types results (New detached homes, used home purchases vs new home purchases, etc).  How many townhomes are selling by certain time-based benchmarks ({1,3,6,12} month period perhaps) Now how do these averages compare to newer builds?  In data exploration is there a tipping point in the data to suggest that the region cannot sustain further builds profitably. Is the population growth in line with the # of townhomes built, and do the price & quality points mesh with the population demographics?  Does the data suggest that people are looking for this style of house vs a detached home?

I think there are a lot of factors that will come into play to predict future demand.

**c) How could the contractor use these models to improve their business model?**

The contractor can use these models to answer their main questions.  Where do I build, what do I build, and how profitable can this expenditure be?  By having predictions for various locations & build sets a contractor could use their experience in conjunction with the models to determine the direction in which to proceed. Ideally an optimal solution is given, and then proceeded upon by the contractor.