NO AI USED

1. Predicting Bacterial Property.

Since I have a "Potato computer without having access to a GPU cluster", I will mostly avoid using neural networks. Now Coming to the procedure and the background as given above, I will use a simple or slightly complex Machine Learning model as describled later. First of all, coming to the dataset, it appears huge for my Potato computer and since we are talking about bacteria with 150 properties, I believe that atleast some of them will be redundant. Therefore, the first task that I will perform is to reduce the dimensionality of the dataset. I have many options here such as PCA which would preserve the variance and project it to lower dimensions or go for Isomaps that would preserve the geodesic dist between neighbours or go for LDA that would create hyperplane on the most discriminative axes. I believe that going with LDA (Linear Discriminant Analysis) would be the best as it would also adap to the Shapes of the complex hyperplanes as well as "remove" the redundant axes. I also suggest that the algorithm can be adjusted to medium scaled datasets like this one by introducing the Schotastic - mini-batch modification. So that the algorithm will be faster as well as it would be not overfitting (in terms of reducing the dimensions. I believe that this can be introduced into a pipeline with a regression

model to prevent overfitting, we can use GridSearchCV
As a regressor, I would prefer polynomial regression due to its ability to fit large dimensional data, but the pipeline of it will contain Stcholastic Gradient Descent regressor rather than the Simple linear regressor due to its ability to behave faster on medium datasets due to its randomised nature. Moreover, the fear of overfitting will also be counterd if we increase the number of iterations of the regressor. I will check for overfitting from the crossvalidation technique from GridSearchCV as it is verymuch optimised towards these operations.

4. OPTIMIZING speed and accuracy in matrix multiplication.

(a) 1. Citations:- A Framework for Practical Parallel Fast Matrix Multiplication.
Austin R. Benson and Grey Ballard
cs. cornell-edu.

To optimize this process, we can use parallelization that is to divide the matrix multiplication task into smaller subtasks & performing them simultaneously on multiple processing units. We can significantly ~~reduce~~ increase the speed by this method.

(b) 2. Citation :- International Journal of Networking & Parallel Computing, Nov 2012, Parallel ~~Algo for~~ Algorithm for Matrix Multiplication

If we want to optimize the algo for accuracy, then, the algo. suggested is Strassen Algo, based on divide & Conquer algo. It reduces the no. of multiplications req. for multiplication & gives more accurate results; We can also use the SIMD processor in O(nlogn) times, ~~but~~ and that too in parallel computing method with multiple GPUs to significantly increase accuracy (by reducing no. of multiplication) and thereby decreasing the total time required to execute the multiplication- This can be ~~pytorch~~ achieved with the help of pytorch or tensorflow as required.

NO AI USED

2. Predicting the number of people on beach.

We will approach this problem in a stepwise menner first starting with analysing and cleaning the data. Then, I suggest that the non necessary variabley can be ignored. This can be done by dimensionality reduction using PCA or ISOMAPS. ~~&~~ Then, I will use a model for prediction of the number of tourists. For this, I will use ~~the~~ a regressor [illegible]. based on decision trees. This is because, the people (visitors) will decide whether to come or not to the beach based on a particular threshold of say temperature or ~~[illegible]~~ cloud cover. So, I believe that we can devise a score for every day based on the weather of the day such that the score will vary based on the parameters depending on the threshold. We will fix the parameter threshold based on the training data. I believe that this may give accurate prediction of the future ~~weather~~ prediction of crowd size based on these weather variables. The dectree based regression is expected to be more accurate due to the reason that it is a threshold based regressor which is more relevant in this case compared to linear regressor.

Citations
~~References~~: Decision tree methods: applications for classification & prediction.

Yan yan Song Ying LV
ncbi.nlm.nih.gov