A Smart Diet Advisor

Akshita Sawhney
MT17143

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Introduction

Motivation

The main motivation behind selection of this problem is the alarming need of an efficient diet advisor which can progressively take an individual towards his/her fitness goals without making him/her go through the turmoil of making efficient food choice. The diet advisor will learn from the preference of the user and would come up with healthier recommendations while taking care of the calorific and nutritional intake.

Formal Problem Statement

Proposing a smart diet advisor which takes users to their fitness goals in a healthier way.

Dataset

The dataset was taken from the below locations:
https://www.kaggle.com/openfoodfacts/world-food-facts/home
https://ndb.nal.usda.gov/ndb/search/list?home=true

The datasets consist of various Branded food items and their nutrient values like Protein,Carbohydrates, Vitamins etc.

Data Preprocessing:

- The columns with more than 80% null values are removed
- The leftover null values were replaced with the mean of the columns and
- The data is then scaled between 0-1
- Top ranked 6 features were selected on the basis of nutritional importance[1]

The final dataset is divided into two sets for first predicting the Calorie:
1) The dataset with the Calorie value with 76180 rows and 17 columns.
2) The dataset without the Calorie value with 1710 rows and 17 columns.

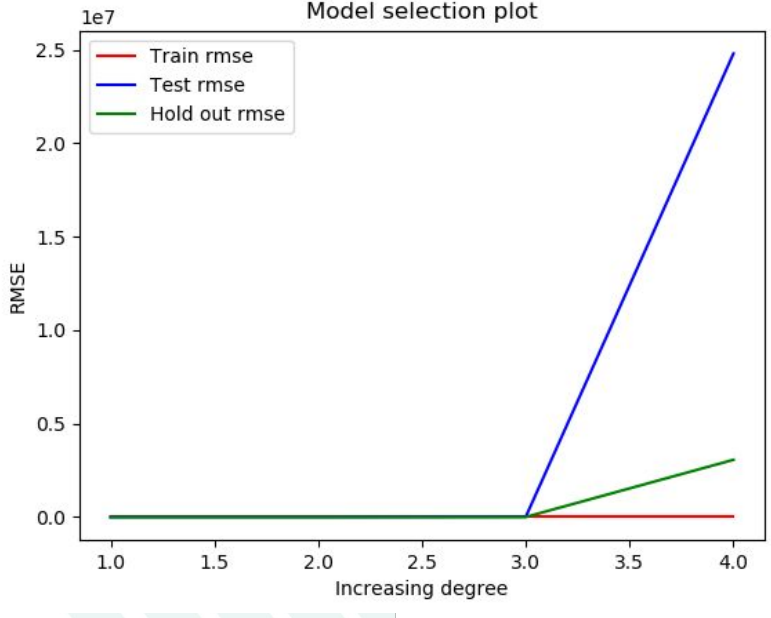
Linear Regression

The first basic approach that we have implemented is **Linear Regression**, where our main motive was to **predict the Calorie value** for all the food items with the value missing. The idea behind this is that for further implementing our recommender system we need the calorie count of the food item.

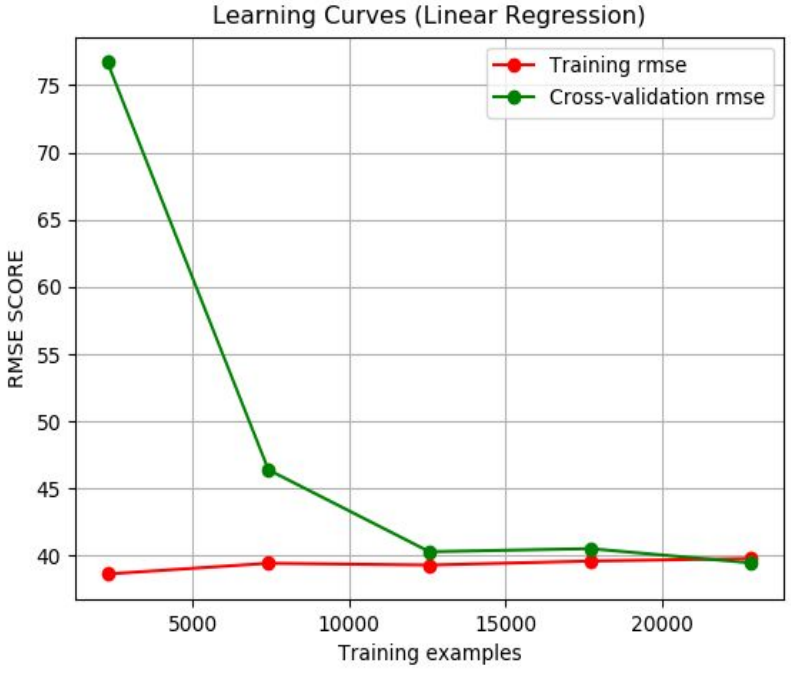
Looking at the graph on the right we observe that with the increasing value of degree , the test rmse and the holdout rmse give a drastic jump in the values, which clearly shows that the model starts to overfit after degree 3.

We chose to stick to degree 1 for polynomial regression, so that we have a simple yet powerful model.
After training the model on the dataset that contained the Energy value we got the below result for 5 fold Cross Validation :

```
Average Train RMSE:  39.66751289173395
Average Test RMSE:  39.57950693789862
Holdout RMSE:  [39.116865351529704]
```



Error Analysis



From the graph on the left we can observe that with increasing training samples the training rmse decreases and also the model becomes more generalized. But after a specific training samples here around 12000, adding more samples does not really help.

```
***** Welcome to Smart Health System, Author: Akshita Sawhney - MT17143 *****
Type 1 for predicting calorific value using trained regressor
Type 2 for using recommendation system built using Self Organizing Maps and K-Means clustering
Type 3 to quit!
1
Please enter protein content per 100grams in grams
0.01
Please enter fat content per 100grams in grams
0.1
Please enter carbs content per 100grams in grams
0.06
Please enter fiber content per 100grams in grams
0.02
Please enter vitamin c content per 100grams in grams
0.00159750079875
121.41245837951172kcal
```

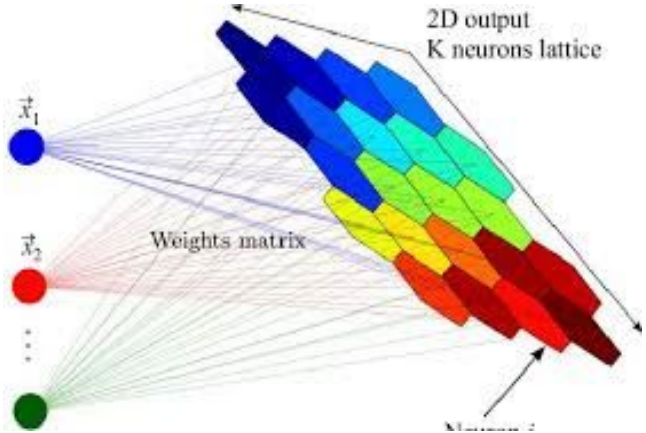
RECOMMENDER SYSTEM: Self Organizing Maps + K means Clustering

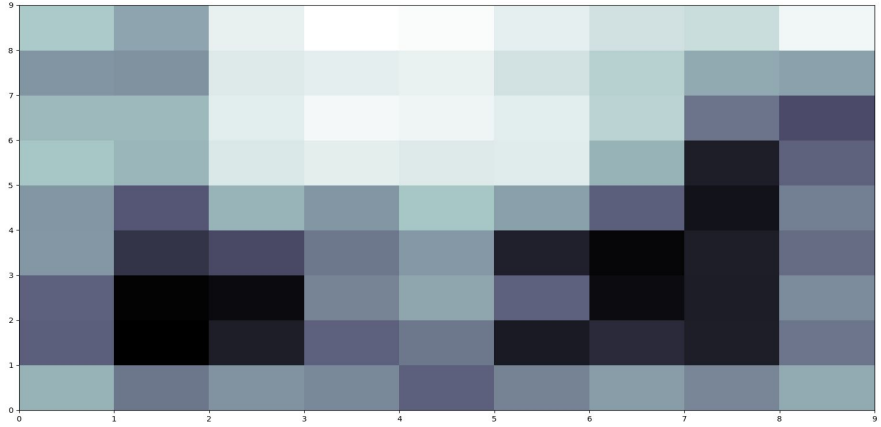
In this system, we used eight nutrients as attributes in the clustering that include carbohydrate, energy, fat, protein, fiber, and vitamin C. All of these nutrients have an influence on the diet in the different ways in both positive and negative.

Our cluster analysis consists of the two stages –

1. Construct and train the SOM, and
2. The SOM is clustered using K-mean approach.

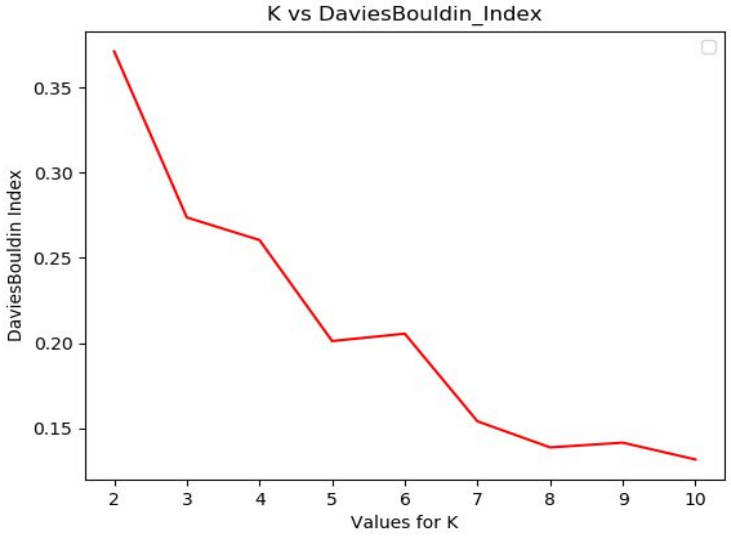
The final result is the food clusters which foods in the same group provide the approximate amount of the eight nutrients.

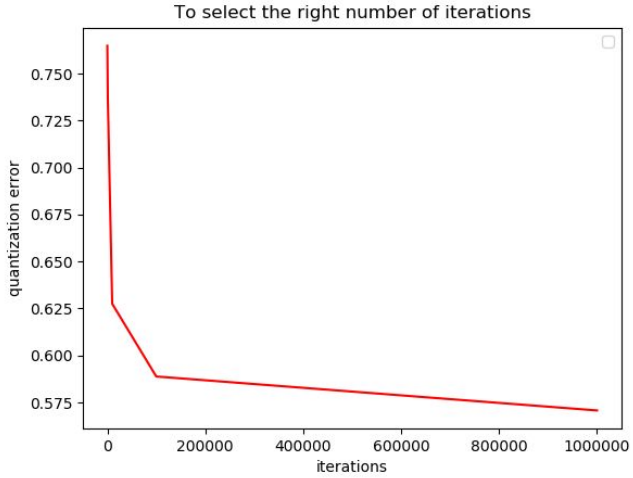




Results

After applying grid search for all parameters, we ended up with quantization error(qe) of 0.2147 for the best parameters. Less the value of qe , better is the quality of som mapping.





Looking at the elbow graph on the top we observe that after k value of 8 the Davies Bouldin Index stays similar. Therefor for K means clustering we choose K = 8 to be the appropriate value.

In the graph above we can observe the quantization error vs iterations graphs which help us decide that for value 100000 the quantization error becomes small and after which it remains similar.

Food Recommender System with UI

```
***** Welcome to Smart Health System, Author: Akshita Sawhney - MT17143 *****
Type 1 for predicting calorific value using trained regressor
Type 2 for using recommendation system built using Self Organizing Maps and K-Means clustering
Type 3 to quit!
2
Please select a food category by typing its category number
Category 0
Category 1
Category 2
Category 3
Category 4
Category 5
Category 6
Category 7
2
Id:45288431, Name:SWEET & HOT DOG BUNS
Id:45288434, Name:SESAME SEED HAMBURGER BUNS
Id:45332199, Name:OLD EL PASO STAND N STUFF TACO DNR KT
Id:45311146, Name:OLD EL PASO STAND N STUFF BOLD TACO DNR KT NACHO CHS
Id:45332699, Name:HIGH PROTEIN ENERGY SHAKE
Id:45334981, Name:LIFE'S BASICS
Id:45327939, Name:FROLLICKS CRNCHY BEAN SEED MIX OLIVE OIL CRACKED BLACK PEPPER
Id:45354355, Name:TOTINOS PIZZA STICKS PEPPERONI
Id:45259345, Name:GARLIC BUTTERY CRUST SANDWICHES
Id:45259334, Name:ARTISAN STYLE MELTS
Id:45064987, Name:HOUSE OF TSANG
Id:45259329, Name:WAFFLES
Id:45206643, Name:HERB OX
Id:45259327, Name:BUTTERMILK PANCAKES
Id:45259323, Name:WAFFLES
Id:45064789, Name:HOUSE OF TSANG
Id:45259322, Name:WHOLE GRAIN WAFFLES
Id:45259317, Name:HOMESTYLE WAFFLES
Id:45286817, Name:ARTISANAL SRIRACHA KETCHUP
Id:45028175, Name:WHOLE WHEAT FLATBREAD WRAPS
Id:45286688, Name:PEACH PIE
Id:45284633, Name:VANILLA BEAN HUMMUS
Id:45269806, Name:PRETZEL CHALLAH
Id:45269745, Name:6 TOMATOES BASIL WRAPS
Id:45269589, Name:ORGANIC PEANUT BUTTER & GRAPE JELLY SANDWICH
Id:45269588, Name:ORGANIC PEANUT BUTTER & STRAWBERRY JELLY SANDWICH
Id:45269373, Name:APRICOT GINGER TERIYAKI
Id:45269181, Name:SOFT DINNER ROLL
Id:45268744, Name:APPLE BERRY CRISP
Id:45268674, Name:GLAZED CROISSANTS
Id:45268487, Name:MEDITERRANEAN BREAD STICKS
Id:45268201, Name:ARTISAN BREAD
Id:45267328, Name:PUMPKIN PIE
Id:45267182, Name:GLAZED DONUTS
Id:45267173, Name:ENRICHED BREAD
Id:45266835, Name:ITALIAN STYLE ICE CREAM GELATO
```

```
Please enter the food id of the selected food item
45216736
Selected item details
Name-BELLA VADO
Energy, Protein, Fat, Carbohydrates, Fiber, Vitamin C
Nutrients-[0.3333333333333333, 0.0, 1.0, 0.0, 0.027909326, 0.000429836]
Calories - unnormalized = 857.0
You have selected a moderately unhealthy item
Our 5 recommendations are as follows:
Id-45163431
Name-EILLIEN'S CANDIES
Energy, Protein, Fat, Carbohydrates, Fiber, Vitamin C
Nutrients-[0.32827693504472966, 0.375, 0.6875, 0.25, 0.094, 0.0]
Calories - unnormalized = 844.0
Id-45145936
Name-AMERICAN VALLEY
Energy, Protein, Fat, Carbohydrates, Fiber, Vitamin C
Nutrients-[0.3306106573317775, 0.35, 0.75, 0.25, 0.1, 0.0]
Calories - unnormalized = 850.0
Id-45254733
Name-FORMAGGIO
Energy, Protein, Fat, Carbohydrates, Fiber, Vitamin C
Nutrients-[0.23609490470633995, 0.6429, 0.6429, 0.1071, 0.0, 0.0]
Calories - unnormalized = 607.0
```