Foundations Of Machine Learning

1) Building a machine learning model that discovers trends in everyday food consumption habits is the task at hand.
Features :
   - Day
   - Breakfast
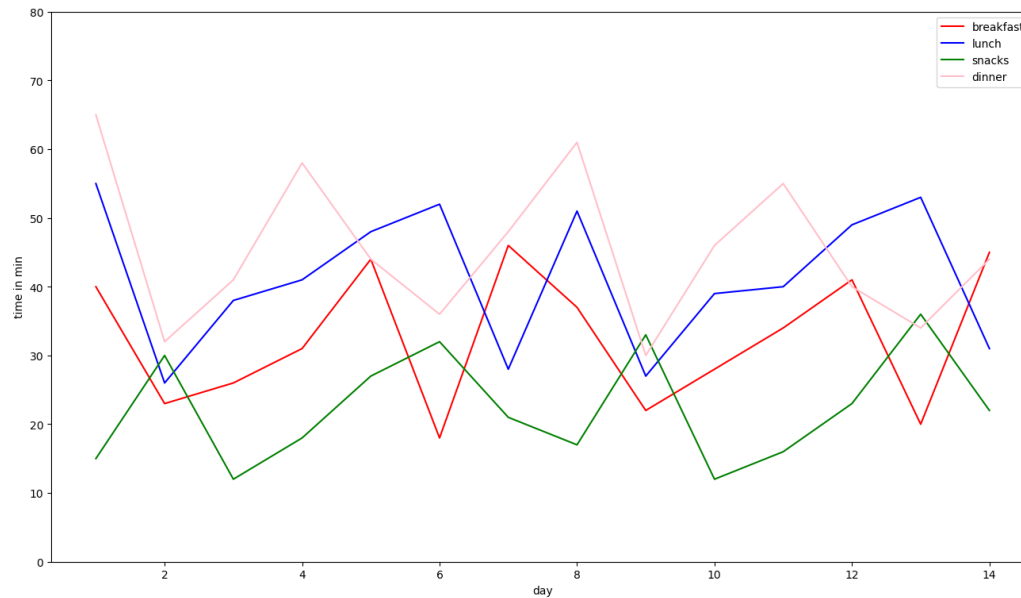   - Lunch
   - Dinner
   - Holiday

   1. Day is the categorical feature which is the basic feature of the model since most of our behaviour (including menu, time table, weekends) depends on the day of the week.
   2. Breakfast feature is 1 if the meal that corresponds to the datapoint is breakfast, else it is 0. The same goes with lunch and dinner. If all three are 0, then the meal of the datapoint is snacks.
   3. Holiday is a binary feature. It takes the value of 1 on holidays and 0 on working days.

   MODEL :
   Poisson regression is a suitable choice when you are dealing with data where the response variable represents counts or events, and you want to model the relationship with predictor variables while considering the inherent variability in the data. In this case, the time taken to finish a meal can be thought of as a count-like variable, since the time can take only discrete integer values in some appropriate range(5-100 minutes) and poisson regression accounts for the variability in this count. Furthermore, it can manage overdispersion, which can happen when meal lengths vary widely.

   DATA :

   The following is a plot of time taken for each meal vs day

From observing the above plot, you can see that

- Snacks take the least amount of time, followed by breakfast, lunch and dinner.
- You can see that in each meal type, days 8 to 14 follow the similar trend as day 1 to 7.
- Meals take a greater amount of time in weekends when compared to weekdays.

The parameters that are learnt by poisson regression model(when optimised by gradient descent) are :

[[2.6981167546665126]
 [0.4809018922909451]
 [0.26601633652645257]
 [0.05133455556930224]
 [0.33053781855645237]
 [0.6332505620786228]
 [0.39705841555812404]
 [0.5390171740866213]
 [0.2468079992559205]
 [0.4852345996019835]
 [0.614909335166498]
 [0.3818723741257487]]

The parameters for breakfast, lunch and dinner(the 9th, 10th and 11th parameters respectively) are in increasing order which is in agreement with the observation 1(time taken order).

In the second to eight parameters, we can see that the fourth parameter is very small, this is to be expected since we encoded day into 7 features, which makes one of them redundant.

Apart from the first parameter which corresponds to constant term, we can see that the parameters corresponding to day have relatively high value. So we can say that the day is the most important feature in our data