# **Q1. Feature Extraction**

Using feature extraction we can try to find out what influences our data. These methods will help us in finding some pattern form the input. For this task I have implemented 4 time series feature extraction methods . I have learnt 2 more tehcniques and couldn’t implement them on the data.

**1. Slope :**

**2. Power Spectral Density (PSD)**

**3. Polynomial fit,**

**4. mean**

**5. Pisarenko PSD using FFT**

**6. FFT**

# **Q2. Why these features**

**1. Slope :**

**The sloper gives us the change in values, so for example if the person has taken food we will have a huge change in the values. This change will let us know that we have food taken by the subject.**

**2. Power Spectral Density:**

Finds us the maximum peak on the power spectrum density, during the meal time. Our input is a Discrete function not continuous , therefore employed discrete fourier transform. And we try to get the frequencies to find the maximum peak ow th power spectrum.

**3. Polynomial fit ( degree 10) :**

**We are trying to fit the curve for a polynomial degree function with degree 10.**

**4. mean :**

**We try to calculate the mean for the series using a sliding window which we can use to learn the difference in means between different time samples. The mean gives us the state where the person will eat**

**Q3. Show values of each of the features and argue that your intuition in step b is validated or disproved? (5 points each ) total 20**

**1. Slope :**

**I have selected slope because my intuition was that whenever the person has food he will have a steep spike on the cgm values. As seen from the figures we can tell that the person has had food.**

**The slope varies from 5-50 and different for every person , So we can conclude that if the person had had food the slope varies in between 5-50. We can also set a threshold that if the value doesn’t cross the minimum then we can conclude that there is no food intake.**

A screenshot of a cell phone

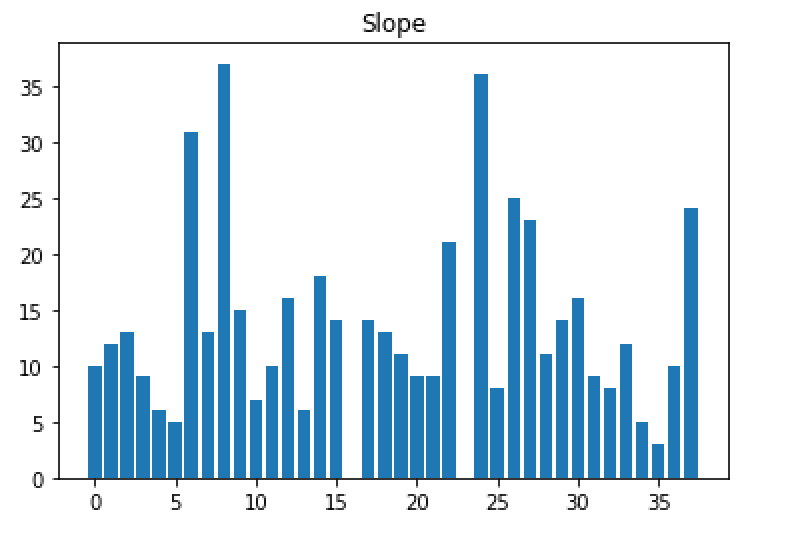
Description automatically generatedA close up of a logo

Description automatically generated

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated



**2. Power Spectral Density :**

The max peak PSD signal represents the meal time .

The max Power spectal density will show us the meal time and the peak values for it. Can also see that the meal peak for the persons varies from 20-35. This feature shows us the maximum peak he has achieved during the meal time. Here too we can set a trhreshold value so as to when does the peak cross the threshold and we can conclude the meal has been taken.

A pencil and paper

Description automatically generated

A pencil and paper

Description automatically generated

A pencil and paper

Description automatically generated

A pencil and paper

Description automatically generatedA pencil and paper

Description automatically generated

**3. Poly fit:**

**I have selected poly fit as another feature because we can use this to fit the values of the input to a curve and we can use the coefficients of the curve as our varibles. As we need higher variance for our features the coefficients would matter because these are the variables that the curve depends on . I have chosen a 10th degree poly fit function such that all the data fit onto to the curve.**

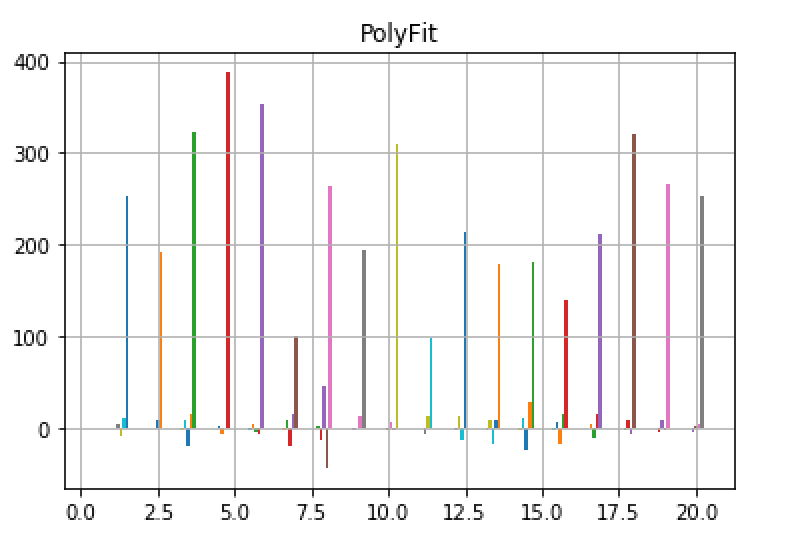
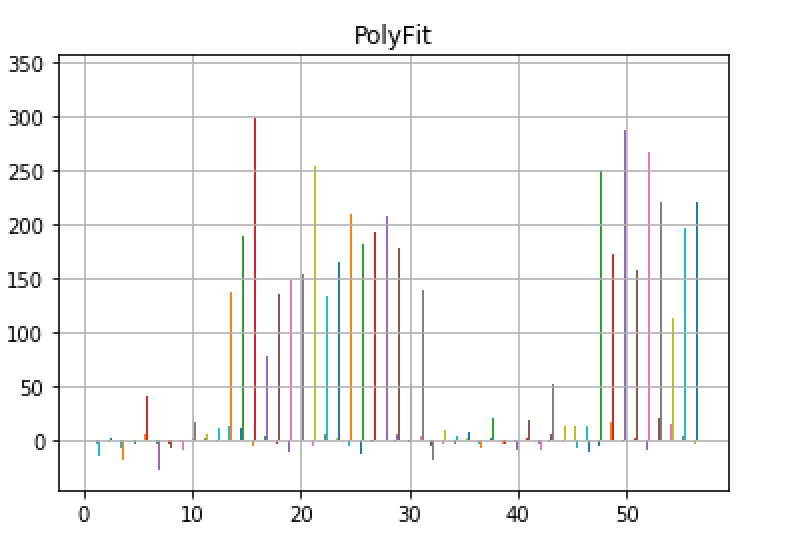
A screenshot of a cell phone

Description automatically generated

A picture containing shoji, object

Description automatically generated

A close up of a logo

Description automatically generated

**4. Sliding Mean:**

**I have chosen this feature because we can know the difference in mean for a sliding window, this change will tell us when has the person consumed food. There will be a spike in the values of the mean. For example if the mean of first 3 is – 20, next 3 is – 35 . Then we know that he has consumed food because of such high variance this feature would be helpful.**

**We can see huge spikes in the data where the mean has changed invariably, this spike would help us know at what value the mean has crossed the threshold. If we take a mean of all these we can actually tell, what at what value the person likely takes in food.**

A picture containing shoji

Description automatically generatedA close up of a logo

Description automatically generatedA picture containing writing implement, pencil, stationary

Description automatically generated

A close up of a logo

Description automatically generated

A picture containing object

Description automatically generated

Q4. Create a feature matrix where each row is a collection of features from each time series. SO if there are 75 time series and your feature length after concatenation of the 4 types of featues is 17 then the feature matrix size will be 75 X 17 (10 points)

Feature Matrix :

Total number of features = 1 slope + Top 5 PSD + 11 coeff of polyfit + length of windowed mean vector.

A close up of text on a white background

Description automatically generated

Q5. Provide this feature matrix to PCA and derive the new feature matrix. Chose the top 5 features and plot them for each time series. (5 points)

A close up of a map

Description automatically generated

# **Q6.For each feature in the top 5 argue why it is chosen as a top five feature in PCA?**

As we know that PCA(ncomponents=5), will chose the top 5 variances in data. AS we give the feature Matrix as the input for the PCA, it’ll will return us the top 5 variances, the eigen vectors gives us the directions where variance will be high. This will return approx. 95% of the variance in the data. We see that feature 1 influences the data more.

# A close up of a map Description automatically generated

# A close up of text on a white background Description automatically generated

A close up of text on a white background

Description automatically generatedA close up of a map

Description automatically generatedA close up of a map

Description automatically generated