Homework Assignment 1

# Question 1

# Question 2

The data set had some initial changes done to it to allow the classifier to use the data much more effectively. Words don’t exactly make any mathematic sense compared to numbers. The 5 different behaviours had their names changed to numbers along with the male female data.

The 5 different behaviours had the following numbers allocated to it:

1 – Sitting Down

2 – Standing Up

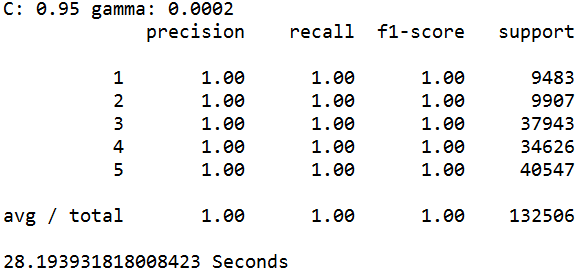
3 – Standing

4 – Walking

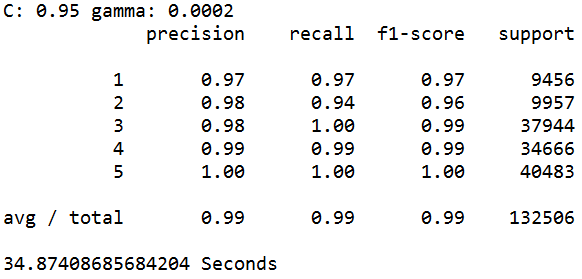
5 – Sitting

The Radial Basis Function Kernel and the Linear Kernel were tested against each other. Both produced exceptional results. After playing with the C and gamma values the RBF Kernel nearly had a perfect score. However, testing the Linear Kernel produced a perfect score. Another thing to note is that since the RBF Kernel makes a nonlinear decision boundary it takes longer to train adding around 25% to the training time. Since the data set consisted of 165000 lines of data entry’s only 20% of the data was used to train the classifier. This is so the classifier isn’t overtrained. Another thing to note is that as the classifier became more accurate the training time also went down by over 50%.

Linear Kernel



RBF Kernel



# Question 3

The initial classifier used on the captcha data was support vector machine. Before attempting to create the classifier, some image pre-processing was done. The images were converted from grey scale to black and white. The increase the size of the data set several things were done. Firstly, the images were shifted left, right, up and down by 1 pixel. Images were also merged to make more variance in the squiggly line crossing the letter.