**Introduction**

This assignment was made by Praneesh Sharma (roll – 21052264) and Mukul Misra (roll – 21052082). This document contains the details of the project which we built as our final assignment in ML Konnexions 2021.

The algorithm that we built here has the purpose of classifying images of rooms as dirty or clean using Logistic Regression, the details of which are elaborated further.

**About the Project**

The algorithm used is a pretty basic one. It is trained using a dataset of images of clean and dirty rooms and uses linear regression to learn to tell the difference between the two.

It does so by getting fed a bundle of test data images, which are grey-scaled and the pixels of which are converted into an array form (using numpy), which makes it easier and faster to analyse. We then send this data to be analysed using logistic regression to build a model capable of classifying any input images of a room as clean or dirty with decent accuracy.

**Algorithm**

The training data set and the test data set of clean and dirty rooms are first converted into an array data structure, which can be easily processed and analysed.

For this, two functions test\_data() and train\_data() are made which open the image, grey-scale it, and resize the pixels to a specific set number. The pixels of the image are then converted into an array form and stored in test\_data and train\_data for the test images and train images respectively, and processed further to be sent off to the Logistic Regression algorithm to form a desired hypothesis.

Logistic Regression separates the data by forming a curve which appropriately distinguishes the images based on the two features. The farther the distance from the curve of a point, the higher will be the confidence of the classification of the data. But since we only want the output in a binary form, the formed hypothesis is passed into the sigmoid function, which converts the distance of a data point from the curve into a probabilistic value of the data being in one class or another. This can be easily used to get a clear output. Gradient descent is then used to optimise the model.

Here, we have used sklearn to make this process easier and clearer.

**Result**

The model we built can predict whether a room is dirty or clean with an accuracy rate of 68.75%.