

Assignment-4

ES20BTECH11015

October 2022

1 KNN

The K - nearest neighbours part of the implementation is done after the pre-processing, The pre-processing uses [pd.get_dummies](#) function to convert the categorical features to binary. After this I split the dataset into three halves for testing, validation and training (132-134-134). I have written my own KNN Model you can see it on my Github

1.1 Results

The results of the Experiments in the [Validation set](#)

$k = 1 \rightarrow 0.8123809523809523$

$k = 3 \rightarrow 0.8095238095238095$

$k = 5 \rightarrow 0.8095238095238095$

$k = 7 \rightarrow 0.7788095238095237$

$k = 9 \rightarrow 0.8171428571428572$

The **best $k = 9$**

2 KDE

Usually making kernels for discrete values is not required, since we do not require any smoothing technique to represent a discrete distribution anyway. You could use a dirac delta Kernel for the discrete values(either individually(on a single feature) or on a random binary vector of features taken from samples). You can use a [Dirac delta kernel](#) for the integer values too, since that distribution is discrete. This part has not been coded up yet.