15ME31001

Piyush Khushlani

Assignment 5

Methodology

Data Preprocessing

Preprocessed the data having different scales and standard deviations by normalizing them with the help of Average and Standard Deviation Vaules stated in the documentation for spam data.

Package Usage

Used C-Support Vector Classification(SVC) of Support Vector Machine under `sklearn` library in Python. By the help of various parameters such as `gamma` (specifying the kernel coefficient) set to `“auto”`, `degree` (for the polynomial functions), `probability` (whether to include the probability estimates) set to `“False”` and `tolerance` (stopping criteria for learning) set to `1e-3`, the accuracies for different `C` (parameter for penalizing error) being analyzed.
Three kernel functions are being tested for Spam Classification with the parameters `gamma`, `probability` and `tolerance` set to their default values as state above:

- `kernel=”linear”` where `degree` is set to 1 by default.
- `kernel="poly"` where `degree` is set to 2 by default.
- `kernel="rbf"` and `degree=3`.

Results

Below are the extremas for the Accuracy Distribution with respect to the corresponding C values

C value	Linear Kernel	C value	Quadratic Kernel	C Value	rbf Kernel
45	0.9217	33	0.9217	1	0.9319
46	0.9217	34	0.9218	2	0.9348
47	0.9218	35	0.9225	3	0.9370
48	0.9218	36	0.9232	4	0.9377
49	**0.9225**	**37**	**0.9232**	**5**	**0.9377**
50	0.9225	38	0.9232	6	0.9348
51	0.9217	39	0.9225	7	0.9355
52	0.9216	40	0.9218	8	0.9355
53	0.9203	41	0.9217	9	0.9341
54	0.9211	42	0.9203	10	0.9348

> *Note: Only those values of C are shown where the Accuracy is reaching its peak value.*