PREFACE

Although support vector machines (SVMs) are one of the oldest machine learning approaches, both the regression and classification variants still help solve real-world tasks. However, the model has a limitation of sensitivity towards noise in the data set. The performance of SVM gets adversely affected when the model is trained with noisy data. This is due to the presence of unbounded hinge loss function. Since the other SVM variants also comprise hinge loss function, the limitation is inherited in all the SVM variants.

In this thesis, the robust formulations of SVM and its variants, which can handle sensitivity towards the noise, are proposed. This comes under the category of robust statistics. For classification tasks, the robustness against label noise in the data sets is added, while the robustness against Gaussian and uniform noise in the data sets is added for regression tasks.

First, the robustness against label noise in the conventional SVM is added. In this literature, the rescaled α-hinge loss function with a non-smooth L_2 regularizer isused. To solve the non-smooth optimization technique, the primal dual proximal (PDProx)-dual technique is implemented. The proposed approach is observed to be more sparse than the existing robust SVMs. The model converges at the rate of $O(1/T)$ where T denotes the number of iterations.

Next, the robustness against the uniform and Gaussian noise in a regression model, twin support vector regression (TSVR) has been added. The rescaled hinge loss func-

tion is used in this literature. To solve the non-convex problem, the half-quadratic optimization technique is used. Subsequently, an algorithm, Res-TSVR, has been devised to implement the proposed approach. It is found that the maximum number of iterations required to achieve an ϵ-precision solution is $O(\log(1/\epsilon))$.

Next, the twin support vector machine (TWSVM) is used in diabetic retinopathy detection using eye fundus images. In this work, the pinball loss function is used with TWSVM to add robustness to the model. It also helped in reducing the computational time. In all the experiments, the proposed approaches are compared with the existing approaches to prove the superiority of the proposed works.

The above contributions discussed are in the field of supervised machine learning. A semi-supervised machine learning framework has also been made robust to the label noise. Next contribution belongs to the addition of robustness to the conventional transductive support vector machine (TSVM). The semi-supervised learning model is robustified against label noise using a truncated pinball loss function. The model is tested on both small and large scale real-world data sets. The proposed approach is implemented using both the stochastic gradient descent (SGD) method and the dual problem solver, mlcv_quadprog (name is based on the machine learning-computer vision lab) . The robust model is also applied to the detection of COVID-19 infected patients using chest X-ray images.