INFO 284 – Machine Learning
Spring 2022

Lab week 8 (Feb 21th – Feb 25th)
Linear classification and SVM

Linear classifiers are popular. The reasons are that logistic regression and linear SVC are fast to train with many data points, and kernelized SVM are good at fitting highly non-linear data.

One of the data sets we have worked with is the churn data set:

https://www.kaggle.com/blastchar/telco-customer-churn

We shall work with this data set also this week.

Tasks:

1. Run classifications on the churn data set with logistic regression, linear SVC, and kernelized SVC.
2. Try to optimize parameters of the learning algorithms by using cross-validation
3. Measure running times for the algorithms
4. Compare with results from Lab 2
5. Try out visualization techniques found in the text book. Some nice examples can be found on pages 56-65 and 95-103.
6. Assess algorithms in terms of accuracy, time spent on learning and also model use, and understandability of final model. Which of the models would you prefer for this data set? What do you think you would prefer if the data set was 1,000,000 data points and not about 7,000?