

COMP 6970: Assignment 5

Participant: Mousumi Akter, Chase Townson, Kushagra Kushagra

Part 1:

Type I Error - This is when the machine learner predicts positive but the actual label is negative. It is also known as a false positive.

Type II Error - This is when the machine learner predicts negative but the actual label is positive. It is also known as a false negative.

Results are average of 10 runs for each MLs

KNN

Predicted	Actual	
	+	-
+	TP(22.8)	FP(13)
-	FN(19.5)	TN(63.7)

SVM with Regression

Predicted	Actual	
	+	-
+	TP(0)	FP(35.8)
-	FN(0)	TN(83.2)

Linear SVM

Predicted	Actual	
	+	-
+	TP(4.4)	FP(31.4)
-	FN(2.2)	TN(81)

MLP

Predicted	Actual	
	+	-
+	TP(16.3)	FP(19.5)
-	FN(9.7)	TN(73.5)

Part 2:

A. KNN

- Identify at least 3 instances of Type I and Type II errors (3 Type I Errors and 3 Type II Errors) for each of the 4 baseline MLs.
 - Result shown for each run in output file (Run[0-9].txt)
- Discover at least 3 instances of Type I and Type II errors that are problems for all 4 baseline MLs (3 Type I errors and 3 Type II errors that are universal errors across all 4 baseline MLs).
 - Result shown for each run in output file (Run[0-9].txt)

Observation:

- SVM with regression performs better to predict negative
- KNN performs better to predict positive
- KNN has highest type-2 error or FN
- SVM with regression has highest type-1 error or FP
- For type-1 error there were common instances that cause universal errors across all 4 baseline MLs for all 10 runs
- For type-2 error there were no common instances that cause universal errors across all 4 baseline MLs for any 10 runs
- After inspecting we saw the dataset were unevenly distributed like more instances with positive labels than negative labels. With less instance with negative label, it's less likely to get instance which causes universal type-2 (FN) error

