Xiaodong Wang xwang322@wisc.edu

CS 760 Machine Learning Homework #1

Data Set information: I obtained the data from my own research based on different elements proportion in semiconductor lasers to determine their emitting wavelength. This data set has 47 examples and based on 11 different attributes.

Attribute Information:

@attribute id number (No. from 1 to 47) numeric

@attribute RI (refractive index) numeric

@attribute As (Arsenide proportion in total) numeric

@attribute In (Indium proportion) numeric

@attribute P (Phosphorus proportion) numeric

@attribute N (Nitrogen proportion) numeric

@attribute Cu (Copper proportion) numeric

@attribute Si (Silicon proportion) numeric

@attribute Ge (Germanium proportion) numeric

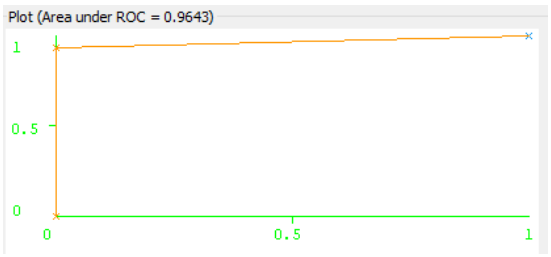
@attribute Ga (Gallium proportion) numeric

@attribute Type of laser (class attribute) numeric
-- 1 visible light -- 2 near-infrared spectrum -- 3 mid-infrared spectrum -- 4 ultra-violent spectrum

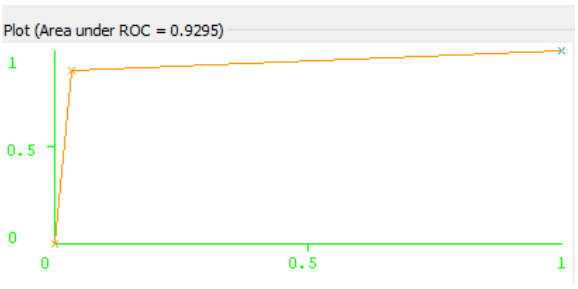
I ran the classifier by both J48 and 1-nearest neighbor and attach the ROC curves (threshold curves) for those two methods (10-fold cross-validation).

J48 ROC threshold curve: (1st: visible; 2nd near; 3rd: mid; 4th: ultra-violent)

ROC area	class
0.9643	visible light
0.9295	near-infrared light
0.9756	mid-infrared light
0.9247	ultra-violent light

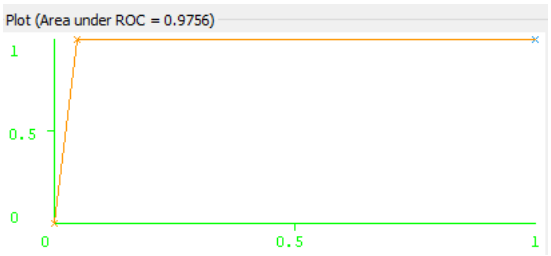


Visible

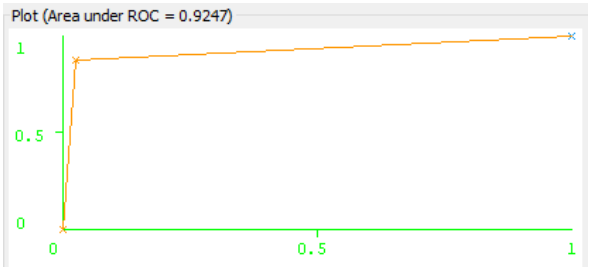


near-IR

Xiaodong Wang xwang322@wisc.edu



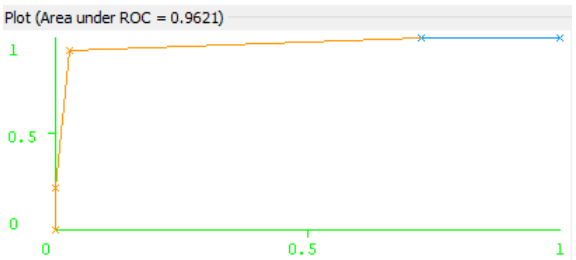
Mid-IR



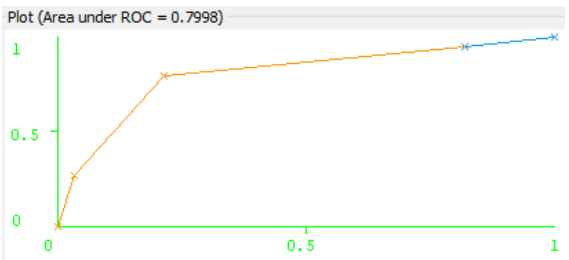
ultra-violent

1NN method ROC threshold curve: (1st: visible; 2nd near; 3rd: mid; 4th: ultra-violent)

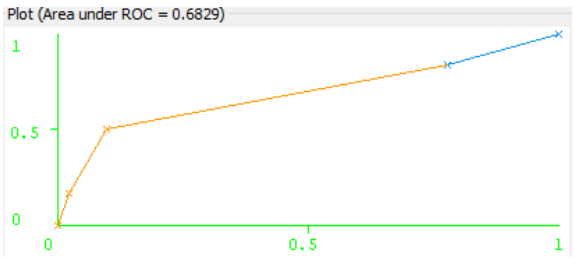
ROC area	class
0.9621	visible light
0.7998	near-infrared light
0.6829	mid-infrared light
0.8654	ultra-violent light



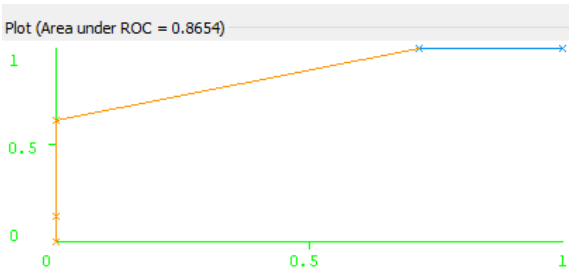
Visible



near-IR



Mid-IR



ultra-violent

Analysis about data set: Semiconductor laser output wavelength is mostly determined by semiconductor materials. Overall the calculation is quite robust, as it can be seen that most of ROC curve is larger than 0.5. The data is more influenced by attribute 6, 7 and 10. Their states would influence laser states a lot when fabricating lasers. And it is known that mid-infrared laser has more Copper and Sulfur than other. That is the reason in one-NN method, ROC is lower than others. As I am studying more concerned in engineering, the data analyzed here is not so uncertain to classify so most of ROC is high than ordinary cases in human science.