

Sistemas Inteligentes

Trabajo 3 – Clasificación Parte 2

Profesor: Alejandro Figueroa

Ayudantes: Alexander Espina - Daniel Palomera

Fecha de publicación: 17/08/15

Fecha de entrega: 11/09/2015

Aspectos generales

- La entrega del informe debe ser realizada, de manera impresa, en la fecha indicada, en la secretaría.
- Se debe utilizar los datos etiquetados por todos los integrantes (máximo dos).

Objetivos

Fortalecer y complementar los conceptos vistos en la tarea anterior, tales como cross-validation y métricas de evaluación. Utilizar un package de herramientas de minería de datos. Comparar diferentes modelos de clasificación.

Desarrollo

El alumno debe comparar el desempeño obtenido por SVM con los siguientes seis modelos: Bayes, Árboles de Decisión, MaxEnt, MaxEntL1, Balanced Winnow y Winnow. Con este propósito, se utilizará la herramienta MALLET:

<http://mallet.cs.umass.edu/index.php>

En un primer paso, el alumno debe importar los splits de entrenamiento calculados en la tarea anterior al formato MALLET. Nótese que para que sus resultados sean legítimamente comparables, debe mantenerse los splits originales. Para importar los archivos debe utilizar la siguiente línea de comando:

```
bin/mallet import-svmlight --input train_file --output train_file.mallet
```

Una vez portados los archivos, se debe generar los diez modelos de entrenamiento, uno por cada split. Para eso, utilice la siguiente línea de comando:

```
bin/mallet train-classifier --input training.mallet --output-classifier my.classifier  
--trainer MaxEnt
```

Para poder hacer ésto con cada uno de los modelos mencionados anteriormente, se debe cambiar el parámetro “MaxEnt” por el nombre del clasificador: Bayes, DecisionTree, BalancedWinnow, MaxEntL1, etc. Es decir, el alumno va a tener por cada split de entrenamiento distintos modelos generados por cada uno de éstos algoritmos de aprendizaje. Posteriormente, para obtener las predicciones mallet utiliza la siguiente línea de comando:

```
bin/mallet classify-file --input data --output - --classifier classifier
```

Con las predicciones obtenidas para los diferentes algoritmos de aprendizaje, se debe comparar sus accuracy, precisión, recall y F-Score. Además, muestre sus curvas ROC. En el caso de tener más de una clase, asuma la clase mayoritaria como la positiva, y el resto negativo. Calcule la curva Lift. Tomando en cuenta el mejor modelo, muestre la matriz de confusión y analice los errores utilizando ejemplos.

Nótese que el informe no sólo contempla el despliegue de los resultados, sino que el alumno debe utilizar una estrategia de presentación de resultados clara y analizar los resultados.