

mÉTODOS DE APRENDIZAJE ESTADÍSTICO, ÁRBOLES DE DECISIÓN Y DISEÑO EXPERIMENTAL

PRÁCTICA 2 – INTELIGENCIA COMPUTACIONAL



antía FRAGA LOURIDO

VERÓNICA ARANDA JARRETA

1. Introducción

En esta práctica se han entrenado varios modelos para dos conjuntos de datos (Iris y Breast Cancer Wisconsin) empleando diferentes métodos para la estimación del error. Los modelos entrenados han sido un modelo con el discriminante lineal de Fisher, otro con el discriminante cuadrático y tres con árboles de decisión. Los dos métodos estadísticos utilizados han sido ‘K-fold cross-validation’ (para el conjunto de datos Breast Cancer Wisconsin) y ‘Leaving one-out cross validation’ (para el conjunto de datos de Iris).

1. Materiales

* Dos conjuntos de datos:
  + **Iris**: 150 muestras, 4 variables y 3 clases.
  + **Breast Cancer Wisconsin (original)**: 699 muestras, 10 variables y 2 clases.
* MATLAB

1. Métodos

En el caso del conjunto de datos de Iris, no ha sido necesario un preprocesado de los datos, sin embargo, este procedimiento si se ha tenido que realizar en el caso del conjunto de datos de Breast Cancer Winsconsin al tener ‘missing values’.

Ante esta situación, se ha decidido sustituir los ‘missing values’ por la media del atributo en el que se encontraban.

Para el conjunto de datos de Iris se ha entrenado:

* Leave one-out:
  + Un modelo empleando el discriminante lineal de Fisher
  + Un modelo empleando el discriminante cuadrático
  + Tres modelos empleando Árboles de Decisión, variando las métricas empleadas en cada uno

Para el conjunto de datos de Breast Cancer Wisconsin:

* K-fold (K = 10):
  + Un modelo empleando el discriminante lineal de Fisher
  + Un modelo empleando el discriminante cuadrático
  + Tres modelos empleando Árboles de Decisión, variando las métricas empleadas en cada uno

Además, para cada modelo entrenado y también para el conjunto de entrenamiento, se calcularon una serie de métricas estadísticas:

* Accuracy (Exactitud)
* Recall (Sensibilidad)
* PPV (Precisión)
* VPN (Valor predictivo negativo)
* Especificidad
* F1 Score

1. Resultados
2. Conclusiones