

Ejercicios de Clasificación de Patrones

Abril 2005

Con los datos del problema de desambiguación de *presente* o los de *conversión grafema-fonema del Holandés* que podréis encontrar en <http://www.ii.uam.es/~jporta/iula>:

1. Comparar los distintos métodos de evaluación: con el *training-set*, con *cross-validation* (variando el número de *folds*) y con *Hold-out* (variando en porcentaje de *split*).
2. Observar los tiempos de aprendizaje de distintos métodos y hacer una comparativa con los resultados y el error de cada método.
3. Para el método de vecinos próximos, encontrar la k óptima para el problema.
4. Interpretar gráficamente los diagramas de dispersión del problema. Sacar conclusiones sobre correlación entre atributos.
5. Variar el número de atributos (eliminando algunos) del problema y resumir los resultados en una tabla. ¿Qué atributos son los más relevantes?
6. Obtener, por distintos métodos, reglas de decisión y observar su significación lingüística.
7. Comprobar la superioridad o no de *bagging/boosting* vs un árbol de decisión.
8. Aplicar *data cleansing* a los datos y volver a entrenar con los nuevos datos obtenidos. Comparar el error en los dos conjuntos de datos.
9. Aplicar árboles de decisión con distintos parámetros para la poda y observar el error y el tamaño de los árboles.
10. Cambiar los valores '??' en el problema de *presente* por desconocidos (-) y observar las diferencias en los resultados aplicando varios métodos de aprendizaje.
11. Cambiando el valor de la semilla en las opciones de evaluación, observar si algún método es inestable.