



Minería de datos

Ejercicios para la comparación de dos métodos de aprendizaje sobre un conjunto de datos



- Conjuntos de datos: Soybean
 - 683 instancias
 - 36 atributos (35 + clase)
 - 19 clases
- Test de McNemar
 - Algoritmos: J4.8, SVM
- Validación cruzada y test de Student pareado
 - Algoritmos J4.8, IB1, SVM, NB
- Validación cruzada y test de Student pareado corregido
 - Algoritmos J4.8, IB1, SVM, NB



Test McNemar:

- Realizar el Test de McNemar, con Weka Explorer
- Generar explícitamente dos particiones disjuntas:
 - 2/3 entrenamiento, 1/3 prueba
 - Filtro no supervisado de instancias: *resample*, opción sin reemplazamiento (prueba: opción invert)
 - Crear el clasificador con conjunto entrenamiento
 - Determinar n_{01} y n_{10} sobre el conjunto de prueba
 - Opción *output predictions* al evaluar
- Aplicar el test sobre J4.8 y SVM



Test de student: validación cruzada sin repetición

- **Validación cruzada con test de Student remuestreado pareado (corregido)**
- Configurar WEKA Experimenter
 - Un conjunto de datos: Soybean
 - Cuatro algoritmos: J4.8, IB1, SVM, NB
 - Validación cruzada 10 particiones, sin repetir
- Realizar Test de Student pareado
- Realizar Test de Student pareado corregido
- En ambos casos, utilizar como algoritmo base primero NB y luego SVM



Test de Student: validación cruzada con repetición

- **Validación cruzada con repetición con test de Student remuestreado pareado (corregido)**
- Configurar WEKA Experimenter
 - Un conjunto de datos: Soybean
 - Cuatro algoritmos: J4.8, IB1, SVM, NB
 - Validación cruzada 10 particiones, 10 repeticiones
- Realizar Test de Student pareado
- Realizar Test de Student pareado corregido
- En ambos casos, utilizar como algoritmo base primero NB y luego SVM



Resultados/discusión/conclusiones

- Para los test de Student, elaborar una tabla para cada algoritmo base y conjunto de datos
- Filas: método de evaluación
- Columnas: algoritmos
- Discutir los resultados (incluido McNemar)
- Elaborar las conclusiones