

FUNDAMENTOS DE APRENDIZAJE AUTOMÁTICO

PRÁCTICA 1: Naïve Bayes y validación cruzada.

1. Resultados de la clasificación

Una vez implementados los distintos clasificadores y técnicas de validación se ejecutaron varias pruebas para clasificar cuatro conjuntos de datos:

- **Car:** Coches clasificados según su accesibilidad
- **Lenses:** Lentes de contacto clasificadas según síntomas
- **Iris:** Flores clasificadas según el tamaño de sus pétalos
- **Spambase:** Emails clasificados como *correo deseado* o *no deseado*.

A continuación se muestra una tabla con el error promedio de dichas pruebas, tomado tras clasificar 5 particiones distintas:

	Naïve Bayes	Naïve Bayes con Corrección de Laplace
Car	0.2759720139930035	0.2690204897551224
Lenses	0.2	0.15
Iris	0.05333333333333333	0.05333333333333333
Spambase	0.36861067837416794	0.36861067837416794

2. Análisis de resultados

Llama la atención que el clasificador Naïve Bayes funcione tan bien con el conjunto *Iris* (con un error del 5% o menos según las particiones). Esto se debe a que los datos se ajustan bien a una distribución normal y a que están muy poco dispersos, ya que tienen una varianza pequeña.

En cuanto a las pruebas en las que afecta la corrección de Laplace, la clasificación mejora, sobre todo en el conjunto de las lentes porque tiene muy pocos datos de entrenamiento.

Como se puede observar, la corrección de Laplace no se aplica sobre los conjuntos que sólo contienen datos continuos (*Iris* y *Spambase*), en su lugar sigue aplicando la función de densidad de la distribución normal.

3. Mediciones sobre el conjunto “Lenses”

Para el conjunto Lenses se tomaron los siguientes valores (sólo para la primera partición, extraídos desde el debugger):

- Probabilidades a priori: $P(\text{Clase}=1)$, $P(\text{Clase}=2)$ y $P(\text{Clase}=3)$:

$$P(\text{Clase}=1) = 0.15$$

$$P(\text{Clase}=2) = 0.2$$

$$P(\text{Clase}=3) = 0.65$$

- Probabilidad de máxima verosimilitud: $P(\text{Tear}=2|\text{Clase}=3)$:

$$P(\text{Tear}=2|\text{Clase}=3) = 0,2307692307692308$$

- La anterior con la corrección de Laplace:

$$P(\text{Tear}=2|\text{Clase}=3) = 0,3076923076923077$$

4. Mediciones sobre el conjunto “Iris”

Para el conjunto Iris se tomaron los siguientes valores (sólo para la primera partición, extraídos desde el debugger):

- Probabilidades a priori: $P(\text{Clase}=\text{Iris-Virginica})$, $P(\text{Clase}=\text{Iris-Setosa})$ y $P(\text{Clase}=\text{Iris-Versicolor})$:

$$P(\text{Clase}=\text{Iris-Virginica}) = 0.4166666666666667$$

$$P(\text{Clase}=\text{Iris-Setosa}) = 0.16666666666666666$$

$$P(\text{Clase}=\text{Iris-Versicolor}) = 0.4166666666666667$$

- Media y desviación típica del atributo continuo PLength condicionado al valor de la clase Iris-Versicolor:

$$\text{Media: } 1.45$$

$$\text{Varianza: } 0.02349$$

$$\text{Desviación típica: } 0.153264$$