UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO
CENTRO UNIVERSITARIO UAEM ZUMPANGO

INGENIERÍA EN COMPUTACIÓN

RECONOCIMIENTO DE PATRONES

LABORATORIO CLASIFICACIÓN CON EL MÉTODO DE NAIVE BAYES

Dr. Asdrúbal López Chau

El método de clasificación Naive Bayes está basado en propiedades básicas de la probabilidad, en específico, el teorema de Bayes. En este laboratorio, implementarás completamente este método de clasificación.

a) Atributos numéricos.

1. Carga el conjunto de datos ***Raisin.csv*** proporcinado con los archivos de este laboratorio.
2. Divide aleaoriamente el conjunto de datos subconuntos, una porción del 20% para pruebas y el resto para entrenamiento.
3. Usando como esqueleto el código orientado a objetos NaiveBayesReal incluido en este laboratorio, implementa los siguientes:
    a. Método fit(self, X, y): Calcula las probabilidades apriori de cada clase y las probabilidades condicionales necesarias, tal como se explicó en clases.
    b. Método predict(self, X): Este método recibe como argumento X, que puede representar a una sola instancia o a un grupo de objetos. El tipo de dato de X puede ser una lista (lista de listas en realidad), un DataFrame de pandas o un array de numpy. En cualquiera de los casos anteriores, el método predict regresa un objeto de tipo Series de pandas, que contiene las predicciones de cada objeto que se pasó al método.
    A manera de ejemplo, se muestran algunos posibles casos para X

```
      x0   x1   x2   x3
0    5.1  3.5  1.4  0.2
1    4.9  3.0  1.4  0.2
2    4.7  3.2  1.3  0.2
3    4.6  3.1  1.5  0.2
4    5.0  3.6  1.4  0.2
```

X es un DataFrame de pandas con cinco instancias, el método predict debe regresar cinco etiquetas como una Serie de pandas

```
In [164]: X
Out[164]: [[4.1, 0.5, 2.4, 0.2]]
```

X es una lista con una instancia, el método predict debe de regresar una etiqueta como una Serie de pandas

```
In [168]: X
Out[168]: [[4.1, 0.5, 2.4, 0.2], [2.1, 5.7, 3.3, 5.1]]
```

X es una lista con dos instancias, el método predict debe de regresar dos etiquetas como una Serie de pandas

```
Out[170]:
array([[4.1, 0.5, 2.4, 0.2],
       [2.1, 5.7, 3.3, 5.1]])
```

X es un array de numpy con dos instancias, el método predict debe de regresar dos etiquetas como una Serie de pandas

Consideraciones:

- Para la clase NaiveBayesReal se supone que todos los atributos del conjunto de datos de entrenamiento son números reales, si no es así, entonces el método fit lanzará una excepción.
- El método predict, deberá de comprobar que todos los atributos sean de tipo numérico real, si no es así, entonces el método lanzará una excepción.

c. Siguiendo las buenas prácticas de programación, crea métodos auxiliares para realizar operaciones necesarias para el método predict, es decir, no se recomienda hacer todo el proceso en un solo método.

d. Mide la exactitud del método de clasificación con este conjunto de datos.

b) Atributos categóricos.

1. Carga el conjunto de datos ***CarDataset.csv*** proporcinado en los archivos de este laboratorio.
2. Usando como esqueleto el código orientado a objetos NaiveBayesCategorical incluido en este laboratorio, implementa los siguientes:
   a. Método fit(self, X, y): Calcula las probabilidades apriori de cada clase y las probabilidades condicionales necesarias, tal como se explicó en clases.
   b. Método predict(self, X): realiza las predicciones, considerando lo mismo que en las instrucciones anteriores.

Consideraciones:

- Para la clase NaiveBayesCategorical se supone que todos los atributos del conjunto de datos de entrenamiento son de tipo categórico, si no es así, entonces el método fit lanzará una excepción.
- El método predict, deberá de comprobar que todos los atributos sean de tipo categórico, si no es así, entonces el método lanzará una excepción.
- En caso de encontrar valores numéricos enteros en los datos, estos serán tratados como cadenas de texto.

e. Siguiendo las buenas prácticas de programación, crea métodos auxiliares para realizar operaciones necesarias para el método predict, es decir, no se recomienda hacer todo el proceso en un solo método.

f. Mide la exactitud del método de clasificación con este conjunto de datos.

Rúbrica

1. **Código con encabezados y documentación (comentarios) importantes para entender el código10%**
2. **Funcionamiento correcto e implementación completa: 80%**

3. **Buenas prácticas de programación, orientación a objetos, seguir el estándar de Python para escritura de código (ver por ejemplo la referencia abajo): 10%**

https://ihumai.medium.com/pep8-un-est%C3%A1ndar-para-escribir-c%C3%B3digo-en-python-96b7d44d4db3