

# Redes Bayesianas

1. **Elabore un modelo de bayes ingenuo (naive bayes) utilizando el conjunto de entrenamiento y explique los resultados a los que llega. El experimento debe ser reproducible por lo que debe fijar que los conjuntos de entrenamiento y prueba sean los mismos siempre que se ejecute el código.**

```
Confusion matrix for Naive Bayes
[[360  29   1]
 [  2  41   4]
 [  0   0   1]]
Accuracy:  0.9178082191780822
```

2. **El modelo debe ser de clasificación, use la variable categórica que hizo con el precio de las casas (barata, media y cara) como variable respuesta.**
3. **Utilice el modelo con el conjunto de prueba y determine la eficiencia del algoritmo para clasificar.**

Se puede decir que la eficiencia de este algoritmo de matriz de confusión es bastante bueno, debido al alto score que se obtuvo, el cual como se puede ver en la imagen previa, es de 91.7%.

4. **Haga un análisis de la eficiencia del algoritmo usando una matriz de confusión. Tenga en cuenta la efectividad, donde el algoritmo se equivocó más, donde se equivocó menos y la importancia que tienen los errores.**

```
Confusion matrix for Naive Bayes
[[360  29   1]
 [  2  41   4]
 [  0   0   1]]
Accuracy:  0.9178082191780822
```

Como se puede observar, la primera columna representa las casas baratas, la segunda las caras, y la tercera las medias. Se puede ver que en total 36 casas fueron mal clasificadas, 30 fueron clasificadas entre caras y medias, mientras que 6 fueron clasificadas como baratas y medias.

5. **Analice el modelo. Explique si hay sobreajuste (overfitting) o no.**

Guido Padilla  
Diego Alvarez  
Oscar Paredez

La efectividad del modelo es de un 92% aproximadamente, y como podemos ver en la tabla hay un total de 36 casas clasificadas erróneamente. Esto nos dice que el modelo no está muy bien entrenado, puede que esté tomando únicamente como válidos los datos idénticos a nuestros datos entrenados. Por lo que no logra distinguir si los datos que salen de los rangos ya preestablecidos son confiables o no. En conclusión, tenemos un 8% de overfitting, que, si realizamos los cálculos apropiados las 36 casas clasificadas erróneamente corresponden al 8% restante de la efectividad del modelo.

**6. Haga un modelo usando validación cruzada, compare los resultados de este con los del modelo anterior. ¿Cuál funcionó mejor?**

```
[0.94174757 0.87254902 0.91176471 0.90196078 0.92156863 0.90196078  
0.87254902 0.8627451 0.83333333 0.94117647]  
0.896135541595279
```

Como se puede observar, el score de la validación cruzada KFold fue levemente menor al score de la matriz de confusión. Por lo poco que varían ambos scores, se podría decir que ambos modelos se adaptan bastante bien a los datos.

**7. Compare la eficiencia del algoritmo con el resultado obtenido con el árbol de decisión (el de clasificación). ¿Cuál es mejor para predecir? ¿Cuál se demoró más en procesar?**

El árbol de decisión tuvo una efectividad del 93% contra un 92% del modelo de bayes. El árbol de decisión demoró 1.9 segundos en ejecutarse, en cambio el modelo de bayes demoró únicamente 0.1 segundos. La diferencia en efectividad es mínima pero en tiempo de ejecución es más efectivo utilizar el modelo de bayes, por lo que este modelo para una mayor cantidad de datos es una buena herramienta a utilizar.