PRACTICA 1

Las fdps no son gaussianas por que la azul a partir de cierto valor (0 aproximadamente) se corta y no tiende a menos infinito. La roja termina en -3.17 y 10.11 aproximadamente y, además, sobre el 4 tiene una depresión, cosa que en gauss no se da.

Si, se trata de la misma distribución estadística, pero con menos datos. El clasificador ideal sería uno no lineal ya que es imposible establecer una frontera de decisión rectilínea que separe los resultados.

Nos conviene utilizar el “Supplied Test Set” porque si utilizamos el “Training Set” estaríamos usando los mismos datos que con los que hemos diseñado el decisor por lo que solo estaríamos memorizando. Con el “Cross-Validation” estaríamos comparando los resultados con un decisor desarrollado combinando los distintos archivos con datos y cambiándolos de función sucesivamente.

Es conveniente utilizar los datos de test porque con los datos de trainnig estaríamos memorizando las decisiones, no aprendiendo de ellas.

Datos Training Set

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| K | 1 | 3 | 5 | 9 | 15 | 17 | 21 | 25 | 31 |
| P(acierto) | 87.25 | 89.083 | 89.9167 | 90 | 90.6667 | 90.333 | 89.9167 | 89.833 | 89.75 |

Datos Training Set Pequeño

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| K | 1 | 3 | 5 | 9 | 15 | 17 | 21 | 25 | 31 |
| P(acierto) | 87.33 | 86.667 | 85.333 | 82.25 | 78.75 | 78.083 | 76.667 | 71.9167 | 68.5833 |

El mejor clasificador es el que utiliza más datos de entrenamiento porque es mucho más preciso.

El mejor valor del parámetro k es 15 con los datos de entrenamiento grande y 1 para conjunto de datos de entrenamiento pequeño. Sí que depende del conjunto de datos de entrenamiento que cojamos, ya que utilizando en conjunto obtendremos un aprendizaje más preciso y nuestra P(acierto) será mayor. La mayor P(acierto) = 90.667 es para el valor de K = 15.