

# Práctica 1: Clasificadores generativos

Jorge Sáez Tejedor  
j.saez.2016@alumnos.urjc.es

Máster Oficial de Visión Artificial  
Universidad Rey Juan Carlos  
Campus de Mostoles, España

El objetivo de esta practica es conocer y emplear la biblioteca Sk-Learn para entrenar y testear varios clasificadores con diferentes conjuntos de datos.

Para ello se dispone de un script en Python que genera los datos de test y entrenamiento, a continuación se presentan esos datos a diferentes clasificadores que realizan una búsqueda de parámetros óptimos mediante *cross validation* 5-Fold. Una vez encontrados se entrena un clasificador con esos parámetros y se evalúa su rendimiento.

En este documento se muestran los resultados, en primer lugar, cada uno de los datos de entrenamiento y test representados gráficamente, a continuación la salida de texto del script Python y por ultimo el resultado de cada uno de los clasificadores entrenados con los parámetros óptimos.

## 1 Resultados con Gaussianas 1

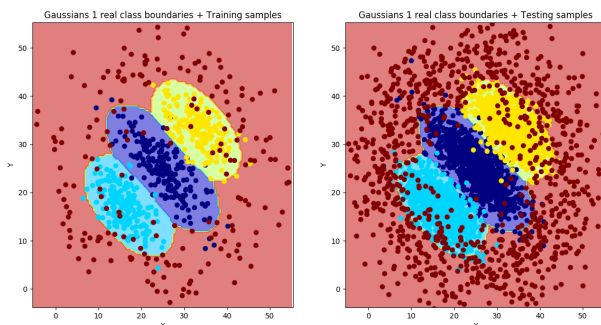


Figura 1: Conjunto de muestra 1

El resultado obtenido para el conjunto de Gaussianas numero 1 es el siguiente:

```
Gaussian Bayes (Test) Done. Score: 0.92
GMM Sk-Learn Done. Score: 0.92 n = [2, 2, 3, 2]
Cov = ['diag', 'tied', 'diag', 'tied']
KNN Done. Score: 0.91 K = 6 Weights = distance
Parzen Done. Score: 0.78 r = 25 Weights = distance
```

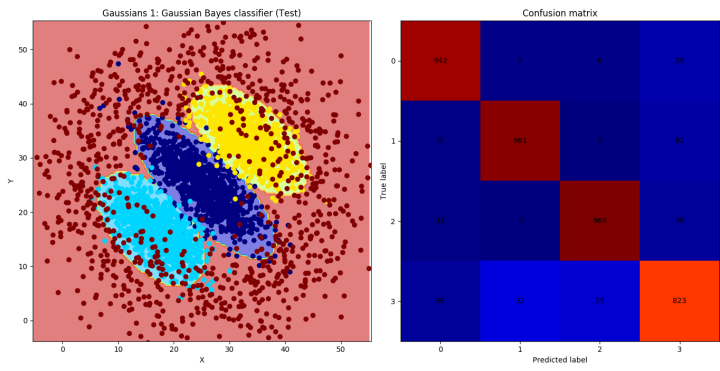


Figura 2: Resultado del conjunto 1 con clasificador de Test Bayesiano

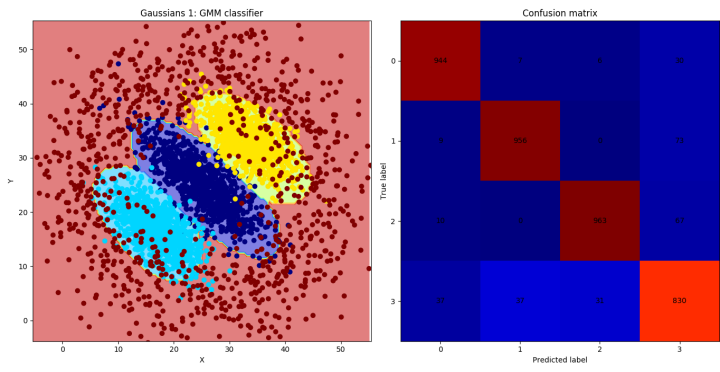


Figura 3: Resultado del conjunto 1 con clasificador de mezcla de gaussianas

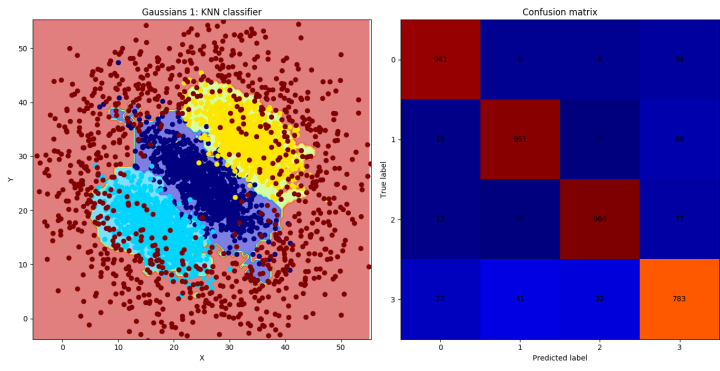


Figura 4: Resultado del conjunto 1 con clasificador de Vecinos mas cercanos (KNN)

A la vista de los resultados, el mejor clasificador para este conjunto de entrenamiento se trata del basado en mezcla de gaussianas, es un resultado mas que esperado ya que,

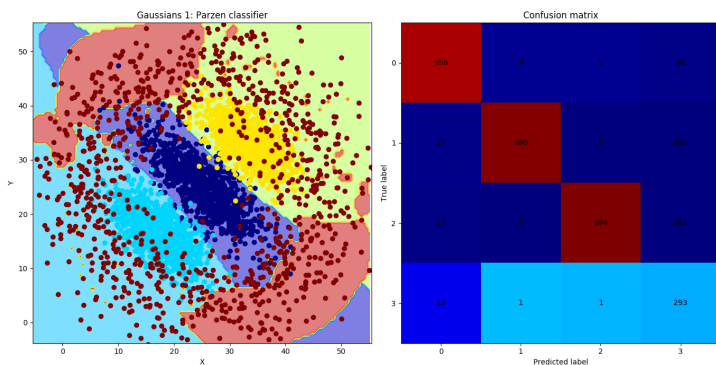


Figura 5: Resultado del conjunto 1 con clasificador basado en ventanas de Parzen

como se puede ver en la Figura 1, 3 de las 4 clases tienen una gran semejanza a gaussianas con matriz de covarianza diagonal, de hecho el clasificador elige este tipo en 2 de las 4.

Como se puede ver en la Figura 5 el clasificador basado en ventanas de Parzen tiene grandes dificultades a la hora de adecuarse a la forma que presentan los datos, sobre todo en las fronteras de la clase 3.

## 2 Resultados con Gaussianas 2

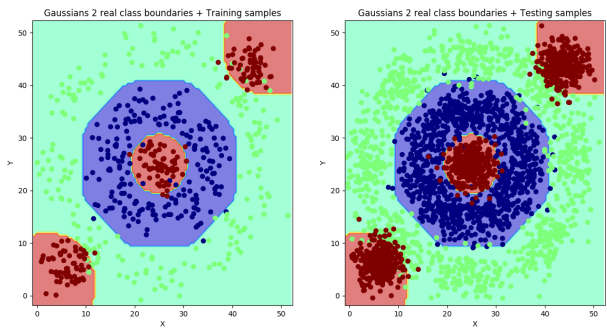


Figura 6: Conjunto de muestra 2

El resultado obtenido para el conjunto de Gaussianas numero 2 es el siguiente:

```
Gaussian Bayes (Test) Done. Score: 0.77
GMM Sk-Learn Done. Score: 0.89 n = [3, 3, 3]
Cov = ['tied', 'spherical', 'tied']
KNN Done. Score: 0.94 K = 4 Weights = distance
Parzen Done. Score: 0.91 r = 52 Weights = distance
```

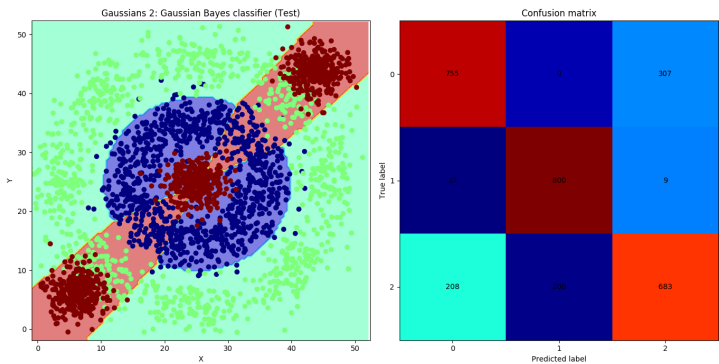


Figura 7: Resultado del conjunto 2 con clasificador de Test Bayesiano

Para este conjunto de datos el clasificador que mejores resultados proporciona es el basado en los vecinos mas cercanos (KNN), el script seleccionó como parámetros óptimos un  $K = 4$  y pesos en función de la distancia. A la vista de los resultados los modelos basados en gaussianas tienen problemas para adaptarse a los datos, principalmente diferenciando entre las clases azul y verde.

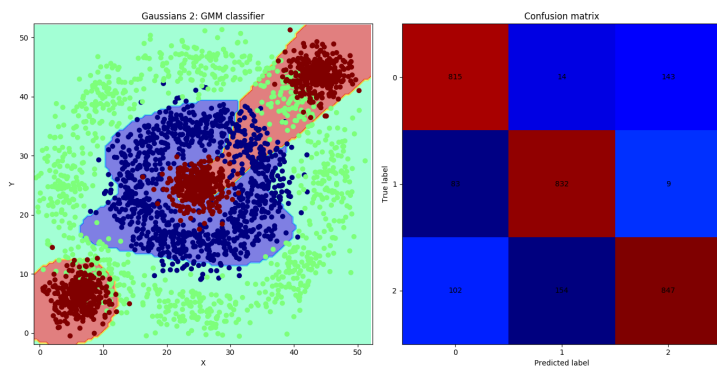


Figura 8: Resultado del conjunto 2 con clasificador de mezcla de gaussianas

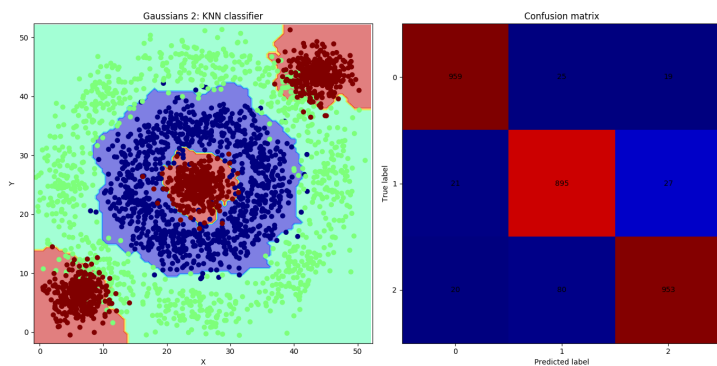


Figura 9: Resultado del conjunto 2 con clasificador de Vecinos mas cercanos (KNN)

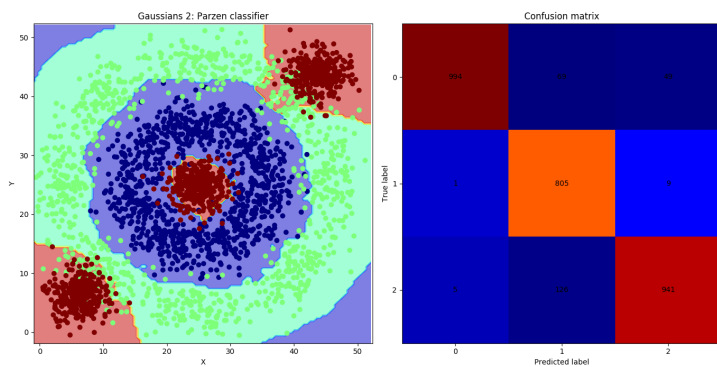


Figura 10: Resultado del conjunto 2 con clasificador basado en ventanas de Parzen

### 3 Resultados con Gaussianas 3

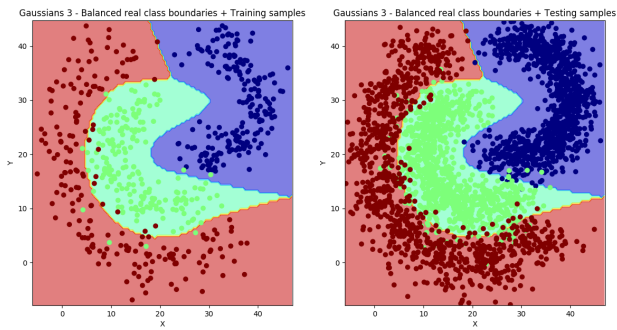


Figura 11: Conjunto de muestra 3

El resultado obtenido para el conjunto de Gaussianas numero 3 es el siguiente:

Gaussian Bayes (Test) Done. Score: 0.93  
GMM Sk-Learn Done. Score: 0.94 n = [3, 2, 2]  
Cov = ['tied', 'tied', 'spherical']  
KNN Done. Score: 0.95 K = 3 Weights = uniform  
Parzen Done. Score: 0.94 r = 49 Weights = distance

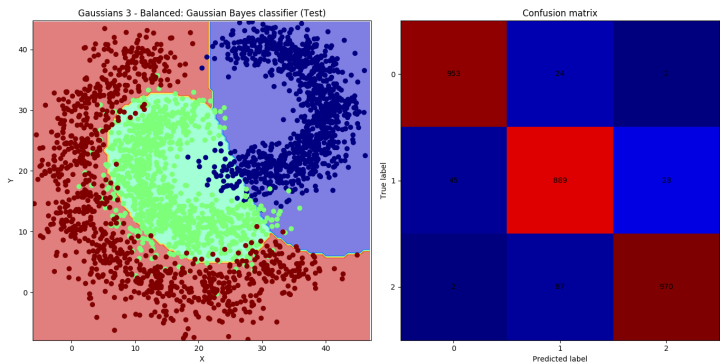


Figura 12: Resultado del conjunto 3 con clasificador de Test Bayesiano

En este conjunto de entrenamiento el clasificador que mejores resultados presenta es KNN, debido principalmente a que los datos presentan datos muy próximos y formas peculiares, a los que los modelos de gaussianas tienen problemas ajustándose

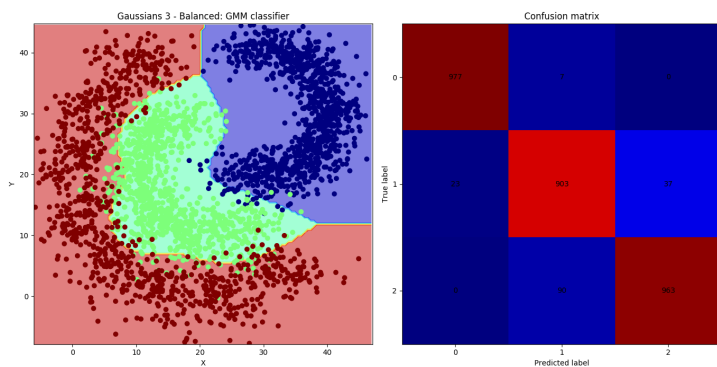


Figura 13: Resultado del conjunto 3 con clasificador de mezcla de gaussianas

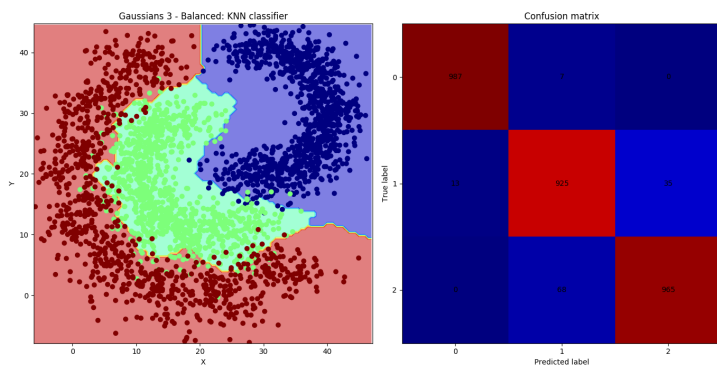


Figura 14: Resultado del conjunto 3 con clasificador de Vecinos mas cercanos (KNN)

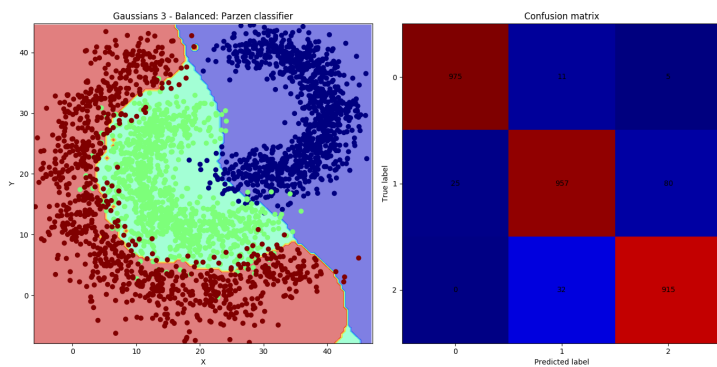


Figura 15: Resultado del conjunto 3 con clasificador basado en ventanas de Parzen

# 4 Resultados con Gaussianas 4

El resultado obtenido para el conjunto de Gaussianas numero 4 (desbalanceado) es el siguiente:

Gaussian Bayes (Test) Done. Score: 0.64  
GMM Sk-Learn Done. Score: 0.96 n = [3, 3, 3]  
Cov = ['spherical', 'diag', 'diag']  
KNN Done. Score: 0.96 K = 12 Weights = uniform  
Parzen Done. Score: 0.64 r = 25 Weights = distance

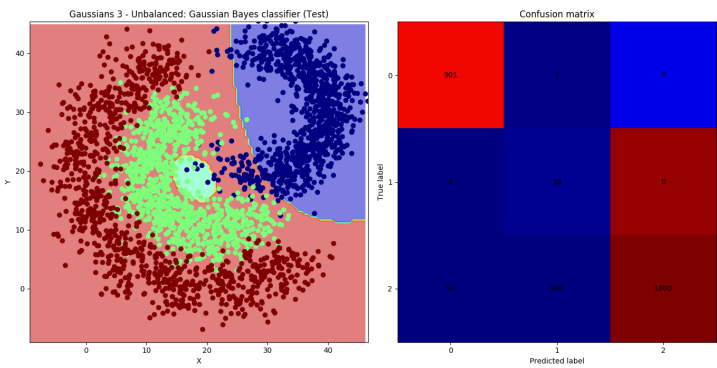


Figura 16: Resultado del conjunto 4 (desbalanceado) con clasificador de Test Bayesiano

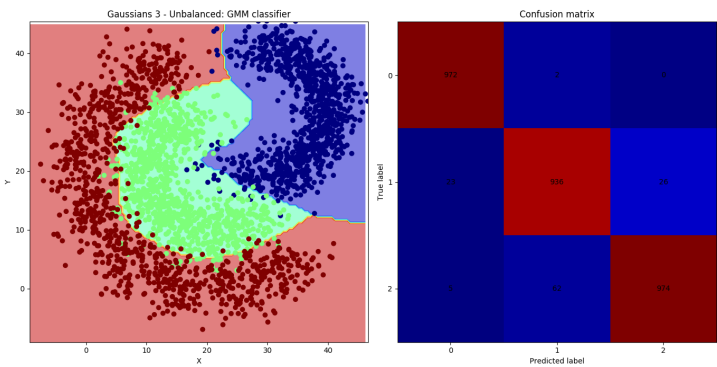


Figura 17: Resultado del conjunto 4 (desbalanceado) con clasificador de mezcla de gaussianas

En este caso las clases presentan diferente numero de datos por clase, lo que repercute en el entrenamiento y el resultado de los clasificadores, para este conjunto el mejor se trata de nuevo de KNN, con un valor mayor de K que en casos anteriores. El basado en mezcla de gaussianas también tiene un buen rendimiento y es capaz de realizar una buena estimación de los datos con tan solo 3 gaussianas por clase.



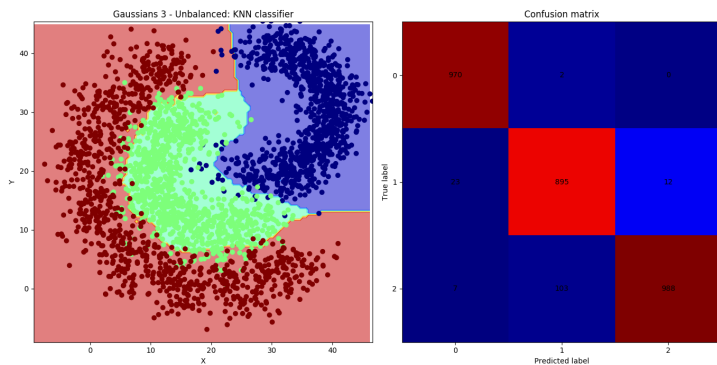


Figura 18: Resultado del conjunto 4 (desbalanceado) con clasificador de Vecinos mas cercanos (KNN)

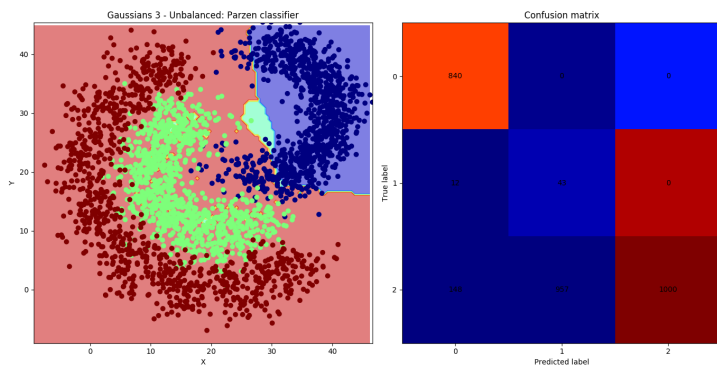


Figura 19: Resultado del conjunto 4 (desbalanceado) con clasificador basado en ventanas de Parzen