

Memoria

MIXTURAS DE GAUSSIANNAS

Mireia Pires State | Aprendizaje automático | 22/11/2022

Contenido

DESCRIPCIÓN DEL TRABAJO REALIZADO	2
RESULTADOS OBTENIDOS.....	2
<i>Con RC = 0.1</i>	2
<i>Con RC = 0.01</i>	3
<i>Con RC = 0.02</i>	3
CONCLUSIONES.....	4

DESCRIPCIÓN DEL TRABAJO REALIZADO

Para el desarrollo de esta práctica se ha usado para la tarea de clasificación, las imágenes de los números del 0 al 9 y sus etiquetas proyectadas mediante PCA a 20 dimensiones.

En cuanto al desarrollo de código, la mayoría ha sido proporcionado por el profesorado. Sin embargo, se han realizado cambios en los archivos para poder generar los modelos dependiendo de la lista de K-componentes que se le pasa al script *mixgaussian-exp.py* cuando lo queremos ejecutar desde el terminal. Un ejemplo sería:

```
python mixgaussian-exp.py train-images-idx3-ubyte.pca20.npz train-labels-idx1-ubyte.pca20.npz '1 2 5 10 20 50 100 200' 90 10
```

Cabe destacar, que los valores escogidos para el estudio son $K = (1, 2, 5, 10, 20, 50, 100, 200)$ y $RC = (0.01, 0.02, 0.1)$. Los últimos valores hacen referencia al factor de regularización de la matriz de covarianza.

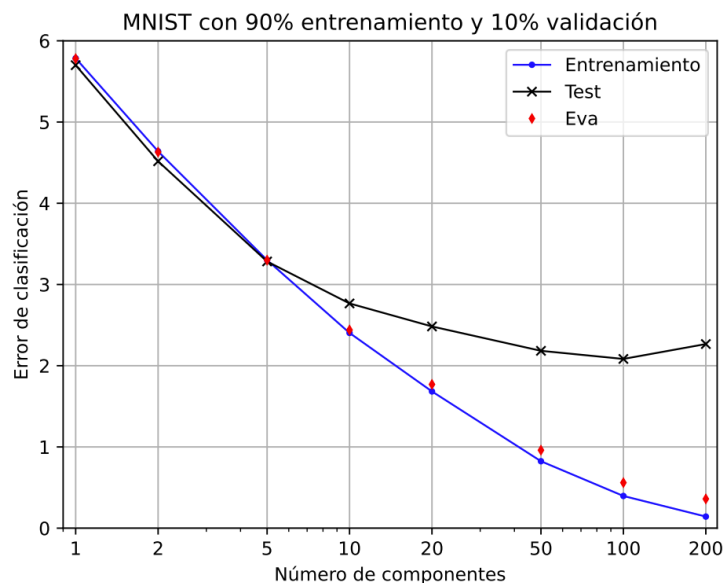
Una vez obtenidos los modelos entrenados con el diferente número de componentes por clase y con un 90% de los datos para entrenarlo, se ha ejecutado el script *mixgaussian-eva.py* para obtener el error del clasificador para todo el conjunto de datos.

Finalmente, ha sido proporcionado un script *mixgaussian-plt.py* por el profesorado que también se ha modificado ligeramente para añadir a la gráfica generada, los resultados obtenidos usando todo el conjunto de datos para hacer el test.

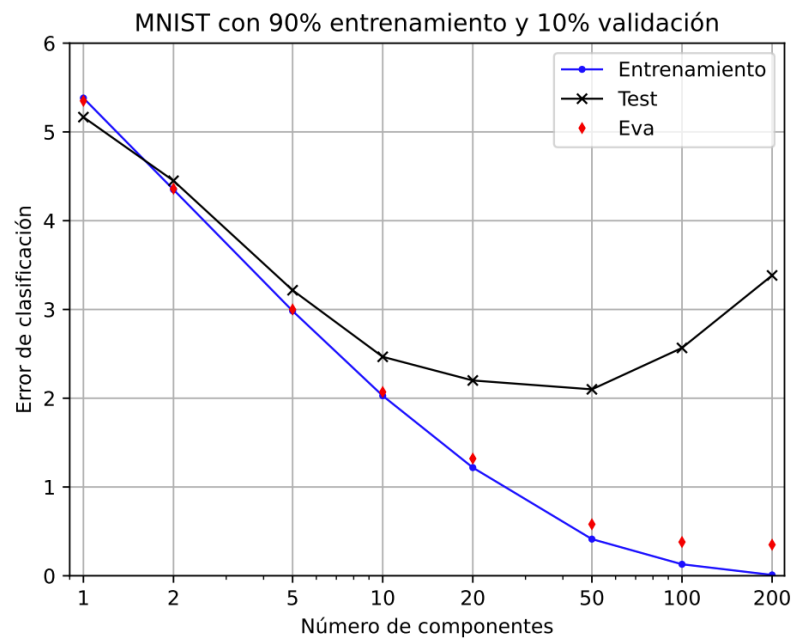
Todos los archivos de código han sido adjuntados a la tarea.

RESULTADOS OBTENIDOS

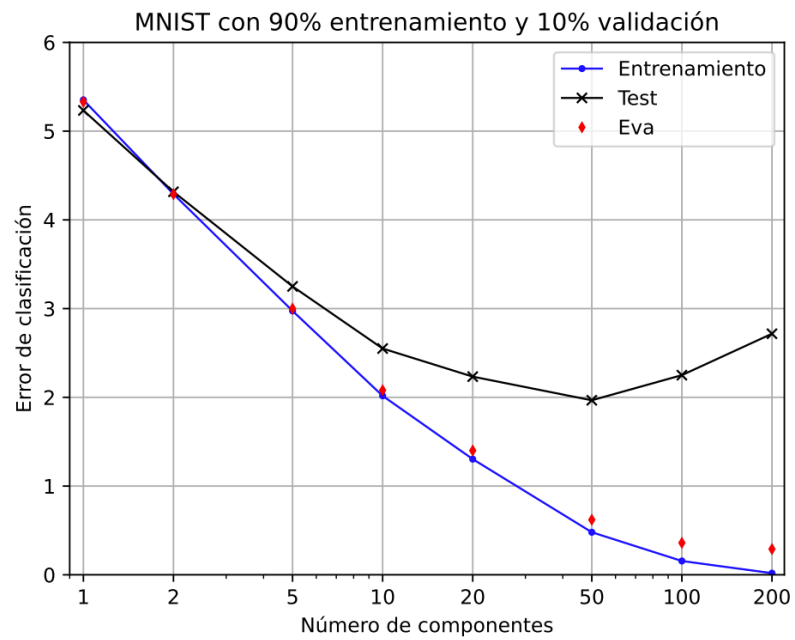
Con $RC = 0.1$



Con $RC = 0.01$



Con $RC = 0.02$



CONCLUSIONES

En primer lugar, en todas las pruebas realizadas se ha detectado que si $K > 10$ componentes, el tamaño del archivo sobrepasa el que se ha pedido (debe de ser menor a 1Mbyte).

Por último, se ha observado que los mejores resultados de entre las pruebas realizadas son con $RC = 0.01$ y $RC = 0.02$ con $K = 10$ ambos.

	Error para $RC = 0.01$	Error para $RC = 0.02$
Entrenamiento	2.03	2.02
Test	2.47	2.55
Eva	2.07	2.08

Como podemos ver para ambos RC los resultados son muy parecidos, pero $RC = 0.01$ da mejores resultados.