

MEMORIA PRÁCTICA 3

Redes neuronales

Ejercicio obligatorio:

Sergi Albiach Caro y Stéphane Díaz-Alejo León (4CO11)

Ejercicio 4.5 (obligatorio: 0.75 puntos). Realiza un experimento donde se evalúe el error de clasificación en función del número de neuronas en la capa oculta de la red neuronal ($n_{\text{Hidden}}=1, 2, 5, 10, 20, 30, 40, 50\dots$), para un número de dimensiones PCA variable ($D = 1, 2, 5, 10, 20, 30$). También puedes utilizar un subconjunto del conjunto de entrenamiento si tienes limitaciones de uso de memoria o si el tiempo de cómputo consideras que es excesivo (i.e. en lugar del 90% usa un 40 %).

Para ello, puedes partir de los scripts `pca+mixgaussian-exp.m` y `mlp-exp.m` para crear el script `pca+mlp-exp.m` que mantenga la misma partición en entrenamiento y validación, y realice al menos la exploración de parámetros propuesta.

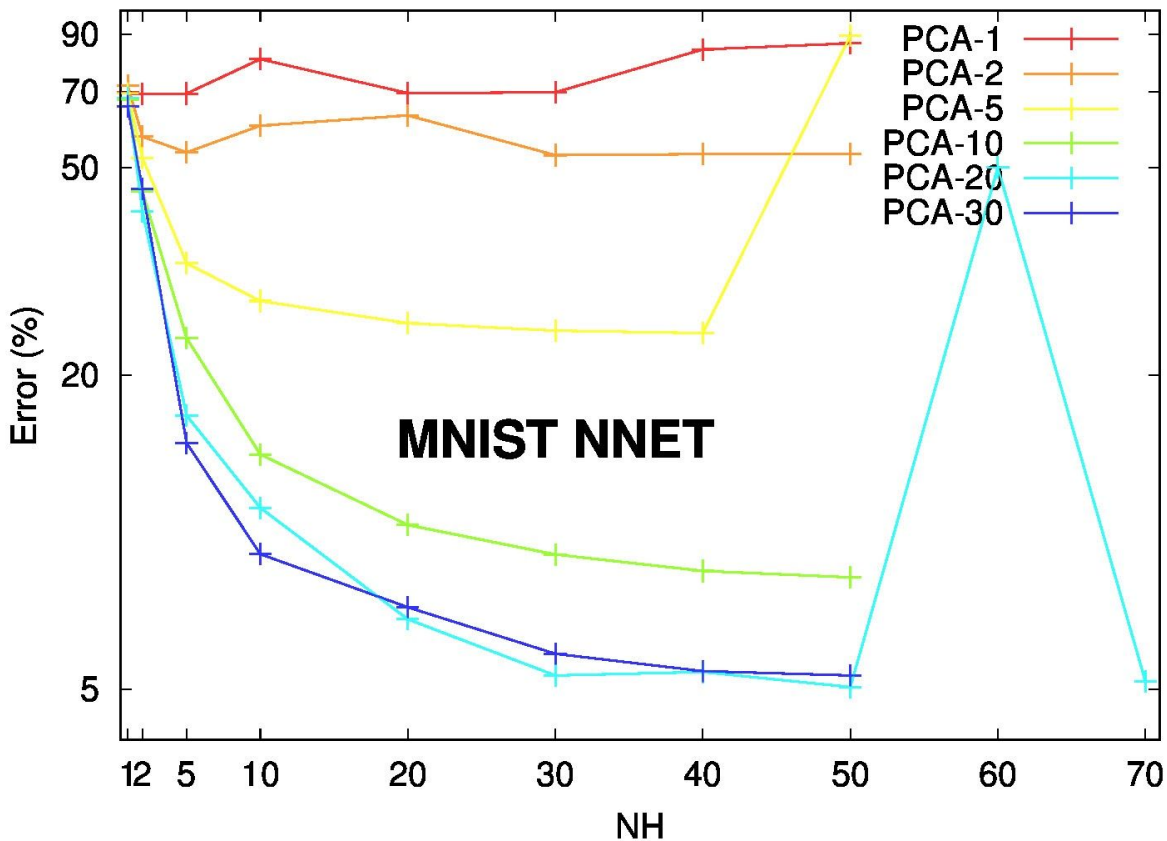
Recuerda que el objetivo de este experimento es determinar los valores de los parámetros del clasificador (n_{Hidden} y D) que minimizan el error de clasificación en el conjunto de validación.

Representa gráficamente las tasas de error obtenidas en el conjunto de validación de forma similar a la utilizada en mezclas de gaussianas. Es decir, cada gráfica tendrá tantas curvas como valores de PCA diferentes hayas probado y cada curva mostrará la evolución del error de clasificación (eje y) en función del número de neuronas en la capa oculta (eje x). Para mejorar la legibilidad de la representación gráfica de los resultados puedes descartar aquellas curvas asociadas a valores de PCA cuyas tasas de error sean muy elevadas.

Debido al enorme consumo de tiempo y memoria los siguientes resultados se obtuvieron con una partición 40/10:

| n_{Hidden}/D | 1 | 2 | 5 | 10 | 20 | 30 |
|-----------------------|--------|--------|--------|--------|--------|--------|
| 1 | 69.700 | 72.000 | 69.550 | 67.617 | 68.150 | 65.700 |
| 2 | 69.233 | 57.383 | 52.200 | 45.017 | 41.317 | 45.517 |
| 5 | 69.450 | 53.550 | 32.800 | 23.583 | 16.750 | 14.833 |
| 10 | 80.783 | 60.267 | 27.767 | 14.100 | 11.150 | 9.083 |
| 20 | 69.533 | 63.050 | 25.183 | 10.333 | 6.817 | 7.183 |
| 30 | 69.800 | 52.900 | 24.367 | 9.067 | 5.317 | 5.850 |
| 40 | 84.350 | 53.150 | 24.100 | 8.433 | 5.400 | 5.417 |
| 50 | 86.633 | 53.183 | 89.617 | 8.200 | 5.050 | 5.317 |
| 60 | - | - | - | - | 50.067 | - |
| 70 | - | - | - | - | 5.183 | - |

La representación gráfica de los resultados es la siguiente:



Se puede observar que conforme aumentan las dimensiones de PCA y de capas ocultas, el error disminuye, alcanzando un mínimo en 20 dimensiones de PCA. Por otro lado, nos hemos encontrado con ciertos valores atípicos en:

- PCA = 5 y nHidden = 50.
- PCA = 20 y nHidden = 60.

Que producen un incremento notable del error. A nuestro parecer, debe producirse debido a que con ese conjunto de parámetros, la red neuronal no aprende como se esperaba.

Para finalizar, los mejores resultados se han obtenido con PCA = 20 y nHidden = 50 dando un porcentaje de error del 5.050%. También cabe destacar que debido a que la partición con la que se lograron los resultados es 40/10, estos resultados pueden que no sean los más precisos, y que con una partición del 90/10 el mejor resultado se encontrase con otro conjunto de valores.

Ejercicio 4.6 (obligatorio: 0.25 puntos). Tras el ajuste de parámetros en el conjunto de validación, en este ejercicio utilizaremos los valores óptimos de los parámetros del clasificador para entrenar y evaluar un clasificador final en los conjuntos oficiales MNIST de entrenamiento y test, respectivamente. Para ello te recomendamos que tomes como punto de partida el script `pca+mlp-exp.m`, modificandolo adecuadamente para generar el script `pca+mlp-eva.m` que también deberá recibir como entrada el conjunto de test de MNIST. Recuerda que toda estimación de (la probabilidad de) error de un clasificador final, debe ir acompañada de sus correspondientes intervalos de confianza al 95 %. Discute los resultados obtenidos comparándolos con los obtenidos en los clasificadores estudiados y con otros clasificadores basados en redes neuronales reportados en la tarea MNIST.

Al principio tratamos de ejecutar esta parte de la práctica con una partición 90/10, pero Octave mataba el proceso, por lo que tuvimos que volver a utilizar una partición 40/10. No obstante, sí que obtuvimos un valor que nos deja entrever que el error descendería, más o menos, 1.2%. Los resultados son los siguientes:

| Partición | D | nHidden | Error | Inter. Confianza |
|-----------|----|---------|-------|------------------|
| 90/10 | 20 | 40 | 3.740 | [3.322 4.158] |
| 40/10 | 20 | 40 | 5.000 | [4.582 5.418] |
| 40/10 | 20 | 50 | 4.770 | [4.352 5.188] |
| 40/10 | 20 | 60 | 4.780 | [4.362 5.198] |

El mejor resultado se sigue obteniendo con PCA = 20 y nHidden = 50. Podemos observar que el error se acerca bastante a los provistos en la web de MNIST:

| | |
|---|-----|
| 2-layer NN, 300 hidden units, mean square error | 4.7 |
| 2-layer NN, 300 HU, MSE, [distortions] | 3.6 |
| 2-layer NN, 1000 hidden units | 4.5 |
| 2-layer NN, 1000 HU, [distortions] | 3.8 |

Per que no presenta una ventaja sobre los estudiados en esta asignatura:

| | |
|------------------|------|
| Gaussian Mixture | 1.62 |
| SVM | 1.95 |