

Práctica 3

Aprendizaje Automático

Ignacio Aguilera Martos

20 de Mayo de 2019

Índice

1. Problema optdigits: clasificación	2
1.1. Problema a resolver	2
1.2. Preprocesado de los datos	2
1.3. Selección de clases de funciones	2
1.4. Conjuntos de training, test y validación	2
1.5. Regularización, necesidad e implementación	2
1.6. Modelos usados y parámetros empleados	2
1.7. Selección y ajuste del modelo final	2
1.8. Idoneidad de la métrica usada en el ajuste	2
1.9. Estimación de E_{out}	2
1.10. Justificación del modelo y calidad del mismo	2

1. Problema optdigits: clasificación

1.1. Problema a resolver

El problema que debemos resolver consta de un conjunto de datos llamado Optical Recognition of Handwritten digits. Contiene en total 5620 instancias con 64 variables más su correspondiente clase. Las clases son 10 (del 0 al 9) indicando el número con el que se identifica la instancia.

Si leemos la descripción del conjunto de datos vemos que todos los datos son numéricos y que no tenemos ningún valor perdido en ninguna instancia. Esto será útil de cara a realizar preprocesamiento de los datos.

Además se provee al final del fichero de descripción del conjunto de datos cómo acierta un modelo K-NN utilizando k desde 1 a 11 donde se puede observar que el porcentaje de acierto es de más del 97 %. Esta información es muy útil, pues si pensamos en cómo funciona el algoritmo K-NN podemos deducir sin pintar ni representar información del conjunto de datos que los mismos están aglomerados de forma clara en clusters. Esto será también relevante a la hora de probar ciertos algoritmos como perceptrón, pues podemos saber más o menos la estructura del conjunto de datos e intuir que va a funcionar correctamente si la separación de los clusters entre sí es suficiente.

1.2. Preprocesado de los datos

1.3. Selección de clases de funciones

1.4. Conjuntos de training, test y validación

1.5. Regularización, necesidad e implementación

1.6. Modelos usados y parámetros empleados

1.7. Selección y ajuste del modelo final

1.8. Idoneidad de la métrica usada en el ajuste

1.9. Estimación de E_{out}

1.10. Justificación del modelo y calidad del mismo