

Tarea 1

Clasificación con "K vecinos más cercanos" (K-NN)

Introducción al Aprendizaje automático
Nareli Cruz Cortés

Utilizando cualquier lenguaje de programación implemente clasificación usando *K-Nearest Neighbor*. Puede utilizar norma 2 (distancia Euclideana) como medida de distancia. Use la base de datos knn-dataset.zip Esta base de datos es una modificación de lo datos de *Optical Recognition of Handwritten Digits Data Set de la UCI machine learning repository*. Contiene los datos para la identificación de los números 5 y 6 escritos a mano.

Disponible en el drive del curso.

No utilice las librerías (como scikit-learn) que ya realizan KNN y validación cruzada. Si lo desea puede utilizar librerías para la operación con matrices (como numpy u otra). Ver la sección 18.8.1 del libro de Rusell y Norving como material de apoyo.

1 Implementación y entrenamiento

Una vez que tenga implementado el algoritmo de clasificación K-NN, será necesario saber cuántos vecinos deben utilizarse, es decir el número óptimo de vecinos, para ello utilice la estrategia de validación cruzada (*k-fold cross validation*). Los archivos ya se encuentran separados en 10 folds.

Ejemplos de entrenamiento:

- Los datos entrenamiento están divididos en 10 archivos (train-

Data1.csv hasta el trainData10.csv).

- Las etiquetas (respuestas de clasificación correctas) están en los archivos trainLabels1.csv hasta el trainLabels10.csv
- Cada archivo tiene 100 ejemplos.

Para el desempeño del algoritmo use **exactitud**
$$\text{exactitud} = \frac{\text{Numero-ejemplos-bien-clasificados}}{\text{total-de-ejemplos}}.$$

Hacer una gráfica que muestre en el eje de las abscisas los valores de K probados, y en las ordenadas la exactitud promedio.

Reportar el número óptimo de vecinos (K^*), es decir el valor de K con que se obtiene la mejor exactitud de clasificación.

2 Etapa de prueba

Como ya sabemos cuál es el número óptimo de vecinos K^* , ya podemos fijarlo en nuestro algoritmo K-NN y estamos en condiciones de probarlo en el conjunto de prueba.

En la etapa de prueba podemos determinar que tan bien clasifica el algoritmo en datos que nunca ha visto antes, para ello utilizamos un conjunto de ejemplos de prueba:

Ejemplos de prueba:

- Los datos para probar están en los archivos testData.csv y testLabels.csv
- Hay 110 ejemplos para probar.

Reportar la exactitud alcanzada en el conjunto de prueba.

3 Reporte

Presentar en la clase del día 25 de agosto 2021.