

**UNIVERSIDAD POLITÉCNICA DE MADRID**

**ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INDUSTRIALES**

**MÁSTER EN AUTOMÁTICA Y ROBÓTICA**

**RECONOCIMIENTOS DE DÍGITOS ESCRITOS A MANO MEDIANTE DIFERENTES TÉCNICAS DE MACHINE LEARNING**

**ASIGNATURA:**

**INTELIGENCIA ARTIFICIAL APLICADA**

**GRUPO 18**

**AUTORES:**

|  |  |
| --- | --- |
| Álvaro Benito Oliva | M20159 |
| Germán Andrés Di Fonzo Caturegli | M20037 |
| Juan José Jurado Camino | M20039 |

Madrid a 18 de junio de 2021

TABLA DE CONTENIDOS

[2 INTRODUCCIÓN Y OBJETIVOS DEL TRABAJO 3](#_Toc74771845)

[3 PROCESAMIENTO DE CARACTERÍSTICAS 3](#_Toc74771846)

[3.1 REDUCCIÓN DE DIMENSIONALIDAD MEDIANTE EL MÉTODO DE ANÁLISIS DE COMPONENTES PRINCIPALES (PCA) 3](#_Toc74771847)

[3.2 REDUCCIÓN DE DIMENSIONALIDAD MEDIANTE DISCRIMINANTE DE FISHER 3](#_Toc74771848)

[3.3 REDUCCIÓN DE DIMENSIONALIDAD MEDIANTE PCA Y AUTO-ENCODER 3](#_Toc74771849)

[4 CLASIFICADOR K-NN 3](#_Toc74771850)

[5 CLASIFICADOR BAYESIANO 3](#_Toc74771851)

[6 PERCEPTRÓN MULTICAPA (MLP) 3](#_Toc74771852)

[7 MAPA AUTOORGANIZADO (SOM) 4](#_Toc74771853)

[8 COMPARACIÓN DE LAS TÉCNICAS UTILIZADAS 4](#_Toc74771854)

[9 CONCLUSIONES FINALES 4](#_Toc74771855)

[10 REPARTO DE ROLES 4](#_Toc74771856)

[ANEXO. CÓDIGO IMPLEMENTADO EN MATLAB 4](#_Toc74771857)

1. [PROCESAMIENTO DE CARACTERÍSTICAS 4](#_Toc74771858)

[II. CLASIFICADOR K-NN 4](#_Toc74771859)

1. [CLASIFICADOR BAYESIANO 4](#_Toc74771860)

[IV. PERCEPTRÓN MULTICAPA (MLP) 4](#_Toc74771861)

[V. MAPA AUTOORGANIZADO (SOM) 4](#_Toc74771862)

# INTRODUCCIÓN Y OBJETIVOS DEL TRABAJO

Explicar en qué consiste el trabajo, cuáles son los objetivos del mismo, qué es el machine learning y un poco sobre qué se ha hecho anteriormente para resolver el problema de los dígitos. Comentar la estructura de la memoria.

# PROCESAMIENTO DE CARACTERÍSTICAS

Explicar para qué se utiliza la reducción de dimensionalidad de los datos de entrada. Comentar por encima que se va a utilizar primero PCA, después Fisher y después una combinación entre PCA y auto-encoder. Después, en cada aparcado, comentar en qué consiste cada uno de los métodos de reducción por encima y explicar cómo se ha aplicado en este trabajo.

Cuando se trabaja con inteligencia artificial, esta precisa de gran cantidad de datos para obtener su máximo partido y debido a esto se almacenan cantidades inmensas de información. Sin embargo, procesar toda esta información supone un costo de recursos muy alto e incluso mucha de esa información no es relevante para la aplicación u objetivo requerido.

El preprocesado es muy importante para obtener resultados de forma eficiente y rápida. Uno de los procesos que se realiza durante el preprocesado, es la reducción de la dimensionalidad, que consiste en reducir el número de variables a tener en cuenta en una colección de datos. Para entender este proceso, se puede pensar en la gran cantidad de pixeles negros invariantes que se encuentran en todas las imágenes de este proyecto. Estos píxeles se mantienen negros en todas las imágenes, por lo que son aportan información para diferenciar los números entre sí.

Otras razones por las que se recurre a la reducción de dimensionalidad son:

* No siempre el mejor modelo es el que más variables tiene en cuenta. De esta forma se consigue reducir el sobreajuste de un modelo.
* Se mejora el rendimiento computacional, traduciéndose en un ahorro en coste y tiempo.
* Se reduce la complejidad y con ello una compresión más sencilla de los resultados.

## REDUCCIÓN DE DIMENSIONALIDAD MEDIANTE EL MÉTODO DE ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

El método de análisis de componentes principales es un método de transformación lineal, no supervisada. Este proceso permite identificar patrones en los datos basándose únicamente en su correlación. Consiste en encontrar las direcciones de máxima variación de los datos y proyectarlos en un subespacio más reducido que el original.

Al ser no supervisado, no se puede obtener un valor de precisión a priori, aunque para asegurar que no se pierda información importante se pueden escoger tantas direcciones de máxima variación se quiera; teniendo en cuenta su autovalor asociado corresponde al valor de la varianza que se está perdiendo al prescindir de esa dirección.

## REDUCCIÓN DE DIMENSIONALIDAD MEDIANTE DISCRIMINANTE DE FISHER

El discriminante de Fisher es un método de reducción de dimensionalidad lineal y supervisado. En este caso sí se puede obtener un resultado a priori de la cantidad de información que se puede perder desde el punto de vista de la clasificación.

## REDUCCIÓN DE DIMENSIONALIDAD MEDIANTE PCA Y AUTO-ENCODER

# CLASIFICADOR K-NN

Explicar en qué se basa el clasificador k-NN. Comentar cómo se ha utilizado para este trabajo y qué cambios se han ido haciendo para ir mejorando el algoritmo.

# CLASIFICADOR BAYESIANO

Explicar en qué se basa el clasificador k-NN. Comentar cómo se ha utilizado para este trabajo y qué cambios se han ido haciendo para ir mejorando el algoritmo.

# PERCEPTRÓN MULTICAPA (MLP)

Explicar en qué se basa el clasificador k-NN. Comentar cómo se ha utilizado para este trabajo y qué cambios se han ido haciendo para ir mejorando el algoritmo.

# MAPA AUTOORGANIZADO (SOM)

Explicar en qué se basa el clasificador k-NN. Comentar cómo se ha utilizado para este trabajo y qué cambios se han ido haciendo para ir mejorando el algoritmo.

# COMPARACIÓN DE LAS TÉCNICAS UTILIZADAS

Comentar las ventajas y desventajas de cada uno de los métodos empleados estableciendo comparaciones sobre todo fijándose en las matrices de confusión. Utilizar también tablas. También podría venir bien hacer una tabla de Excel donde se comparen las precisiones y los recall de cada clase.

# CONCLUSIONES FINALES

Enumerar principales conclusiones del apartado anterior.

# REPARTO DE ROLES

Indicar coordinador, científico y técnico. Poner porcentajes de trabajo de cada uno (la suma debe ser 300%).

# ANEXO. CÓDIGO IMPLEMENTADO EN MATLAB

## PROCESAMIENTO DE CARACTERÍSTICAS

## CLASIFICADOR K-NN

## CLASIFICADOR BAYESIANO

## PERCEPTRÓN MULTICAPA (MLP)

## MAPA AUTOORGANIZADO (SOM)