

Redes Neuronales Artificiales

Trabajo Práctico 1.

1. Introducción

El objetivo de este trabajo práctico es que utilicen su implementación de los modelos vistos de redes neuronales artificiales para aprendizaje supervisado con datos de dos problemas dados. Se espera que, basándose en estos conjuntos de datos, se encuentren modelos adecuados para los problemas, que no sólo puedan ser entrenados correctamente, sino que también tengan una capacidad aceptable para generalizar con otras instancias de datos.

El desarrollo de estas soluciones deberá ser documentado justificando las decisiones tomadas y explicando los resultados obtenidos.

2. Problemas

Los dos conjuntos de datos pertenecen a problemas reales. Uno puede ser visto como un problema de clasificación, en donde se debe determinar a qué categoría pertenece una entrada, y el otro puede ser visto como un problema de regresión, en donde se deben predecir unos valores de respuesta a partir de una entrada dada.

Los archivos se encuentran en formato CSV sin encabezado y, como los datos pertenecen a problemas reales, es posible que para que puedan ser utilizados de forma correcta deban ser preprocesados. Una porción de los datos fue excluida y reservada para que puedan ser utilizados en una evaluación como datos de testeo.

2.1. Diagnóstico de cáncer de mamas.

Este conjunto de datos contiene los resultados de un examen específico que es utilizado en el diagnóstico de cáncer de mamas. Cada entrada corresponde a los datos obtenidos para distintos pacientes y contiene 10 características provenientes de imágenes digitalizadas de muestras de células. Junto con estas características se encuentra también el diagnóstico final, determinado gracias también a otras pruebas, en donde se indica si la muestra analizada pertenecía a un tumor maligno o benigno.

El conjunto de datos tiene como primera columna el *Diagnóstico* (1=maligno, 0=benigno), y la información en valores reales de los atributos *Radio* (media de la distancia desde el centro a los puntos de perímetro), *Textura* (desviación estándar de los valores en escala de gris), *Perímetro*, *Área*, *Suavidad* (variaciones locales en la longitud del radio), *Compacidad* ($\text{perímetro}^2 / \text{área} - 1$), *Concavidad* (severidad de las porciones cóncavas del contorno), *Puntos cóncavos* (proporción de porciones cóncavas del contorno), *Simetría* y *Dimensión fractal*.

Deberá determinarse si es posible utilizar un modelo de red neuronal artificial para predecir el diagnóstico final utilizando únicamente los datos de este examen.

2.2. Eficiencia energética

Este problema consiste en determinar los requerimientos de carga energética para la calefacción y refrigeración de edificios en función de ciertas características de los mismos. El análisis energético se realizó utilizando edificios de distintas formas que difieren con respecto a la superficie y distribución de las áreas de reflejo, la orientación y otros parámetros.

Cada instancia del conjunto de datos contiene 10 valores correspondientes a las características de un edificio distinto. Los primeros 8 son los atributos *Compacidad Relativa*, *Área de la Superficie Total*, *Área de las Paredes*, *Área del Techo*, *Altura Total*, *Orientación*, *Área de Reflejo Total* y *Distribución del Área de Reflejo*, propios de cada edificio. Los dos últimos corresponden a los 2 valores reales de *Carga de Calefacción* y *Carga de Refrigeración* promedio de mediciones anteriores que representan la cantidad de energía necesaria para realizar una calefacción y refrigeración adecuadas.

El objetivo es determinar si es posible utilizar un modelo de redes neuronales para predecir los dos valores de carga energética a partir de los atributos dados.

3. Detalles de la entrega

La entrega deberá consistir de un informe general o uno por problema, en formato *jupyter-notebook/colab*, con el código completo que demuestre el entrenamiento y la evaluación de los modelos.

Este informe deberá ser breve y conciso. En el mismo se debe describir qué modelo de red neuronal artificial fue adoptado como solución para cada problema, especificando la arquitectura elegida, el método de entrenamiento utilizado, tipo de procesamiento a los datos, y los resultados obtenidos. Es importante también que esté documentado el proceso que condujo a los resultados finales, justificando las decisiones tomadas y las conclusiones a las que se hubieren llegado.

Para el desarrollo de estas soluciones se recomienda utilizar únicamente las librerías *numpy* y *matplotlib*. Adicionalmente se puede recurrir a la librería *requests* que permite cargar los datos desde un origen externo.

```
import requests
r = requests.get('https://...url.../datos.csv')
data = numpy.loadtxt( r.iter_lines(), delimiter=',')
```

La inclusión injustificada de otras librerías puede tener un impacto negativo en la calificación del trabajo.

Se espera también que al volver a ejecutar todas las celdas del informe se obtengan resultados similares a los presentados. En caso de que existan dificultades en la ejecución del código para la evaluación, el trabajo puede llegar a ser rechazado.

El material correspondiente a la entrega deberá enviarse a la dirección:

entregas.redneu@gmail.com

La fecha de entrega será confirmada en la presentación de este trabajo y/o por la lista de correo de la materia.