



Universidad Nacional de Rosario
Facultad de Ciencias Exactas, Ingeniería y Agrimensura
T.U.I.A
Aprendizaje Automático II

Trabajo Práctico 1

Redes Densas y Convolucionales

2024

Autor/es:

Grupo N°	
Nombre y Apellido	N° de Legajo

Corrigió	Calificación

Problema 1

Descripción:

En este problema, se presenta un conjunto de datos que contiene información sobre el rendimiento académico de estudiantes universitarios, así como diversos factores que podrían influir en él. El objetivo es construir un modelo de regresión utilizando redes neuronales para predecir el índice de rendimiento académico de los estudiantes basado en las características proporcionadas.

Dataset:

El dataset proporcionado incluye las siguientes variables para cada estudiante:

- **Hours Studied:** El número total de horas dedicadas al estudio por cada estudiante.
- **Previous Scores:** Las puntuaciones obtenidas por los estudiantes en exámenes previos.
- **Extracurricular Activities:** Si el estudiante participa en actividades extracurriculares (Sí o No).
- **Sleep Hours:** El número promedio de horas de sueño que el estudiante tuvo por día.
- **Sample Question Papers Practiced:** El número de cuestionarios de muestra que el estudiante practicó.

Además, el dataset incluye la variable objetivo:

- **Performance Index:** Un índice que representa el rendimiento académico general de cada estudiante, redondeado al entero más cercano. Este índice varía de 10 a 100, donde valores más altos indican un mejor rendimiento.

Objetivo:

Utilizando el dataset proporcionado, el objetivo es construir un modelo de regresión utilizando redes neuronales que pueda predecir con precisión el índice de rendimiento académico de los estudiantes. Se debe entrenar y evaluar el modelo utilizando técnicas adecuadas de validación y métricas de evaluación de regresión.

Entrega:

La entrega debe incluir:

Código fuente de la solución implementada en Google Colab, que incluya:

- Análisis previo y preprocesamiento del set de datos.
- Definición y entrenamiento del modelo.
- Resultados de la evaluación del modelo, incluyendo métricas de desempeño y visualizaciones relevantes.

Nota: el código debe estar debidamente documentado con comentarios explicativos para que el trabajo sea fácilmente comprensible para otros revisores.

Problema 2

Descripción:

En el siguiente problema, se presenta un conjunto de datos sobre diferentes variedades de frijoles secos. El objetivo es desarrollar un modelo de clasificación que pueda distinguir entre siete variedades registradas de frijoles secos, utilizando características como forma, tamaño, tipo y estructura. Se utilizó un sistema de visión por computadora para obtener imágenes de los granos de frijol y luego se llevó a cabo la segmentación y extracción de características para obtener un total de 16 atributos, incluidas 12 dimensiones y 4 formas.

Dataset:

El dataset proporcionado incluye las siguientes variables para cada paciente:

- **Area (A):** El área de una zona de frijol y el número de píxeles dentro de sus límites.
- **Perimeter (P):** La circunferencia del frijol se define como la longitud de su borde.
- **Major axis length (L):** La distancia entre los extremos de la línea más larga que se puede dibujar desde un frijol.
- **Minor axis length (I):** La línea más larga que se puede dibujar desde el frijol mientras está perpendicular al eje principal.
- **Eccentricity (Ec):** Excentricidad de la elipse que tiene los mismos momentos que la región.
- **Convex area (C):** Número de píxeles en el polígono convexo más pequeño que puede contener el área de una semilla de frijol.
- **Equivalent diameter (Ed):** El diámetro de un círculo que tiene el mismo área que el área de una semilla de frijol.
- **Extent (Ex):** La relación de los píxeles en el cuadro delimitador con respecto al área del frijol.
- **Solidity (S):** También conocida como convexidad. La relación de los píxeles en la envoltura convexa con respecto a los que se encuentran en los frijoles.
- **Compactness (CO):** Mide la redondez de un objeto: Ed/L .
- **ShapeFactor1 (SF1).**
- **ShapeFactor2 (SF2).**
- **ShapeFactor3 (SF3).**
- **ShapeFactor4 (SF4).**
- **Class:** Seker, Barbunya, Bombay, Cali, Dermosan, Horoz y Sira.

Objetivo:

Utilizando el conjunto de datos proporcionado, el objetivo es construir un modelo de clasificación que pueda predecir con precisión la variedad de frijol seco basándose en las características mencionadas anteriormente. Se emplearán técnicas de aprendizaje automático, específicamente redes neuronales, para entrenar y evaluar el modelo.

Además, se requiere que el alumno calcule y agregue al conjunto de datos las siguientes características:

- **Aspect ratio (K):** Relación entre la longitud del eje mayor (L) y la longitud del eje menor (I) .
- **Roundness (R):** Se debe calcular utilizando la fórmula $(4\pi A)/(P^2)$.

Estas nuevas características calculadas se utilizarán como parte del conjunto de datos para el entrenamiento y evaluación del modelo de clasificación.

Entrega:

La entrega debe incluir:

Código fuente de la solución implementada en Google Colab, que incluya:

- Análisis previo y preprocesamiento del set de datos.
- Definición y entrenamiento del modelo.
- Resultados de la evaluación del modelo, incluyendo métricas de desempeño y visualizaciones relevantes.

Nota: el código debe estar debidamente documentado con comentarios explicativos para que el trabajo sea fácilmente comprensible para otros revisores.

Problema 3

Descripción:

En este problema, se proporciona un conjunto de datos que contiene imágenes de escenas naturales de todo el mundo. El objetivo es construir un modelo de clasificación utilizando redes neuronales convolucionales (CNN) para clasificar estas imágenes en una de las seis categorías predefinidas.

Dataset:

El dataset proporcionado contiene alrededor de 25,000 imágenes de tamaño 150x150, distribuidas en seis categorías:

- **buildings**
- **forest**
- **glacier**
- **mountain**
- **sea**
- **street**

Las imágenes están divididas en tres conjuntos:

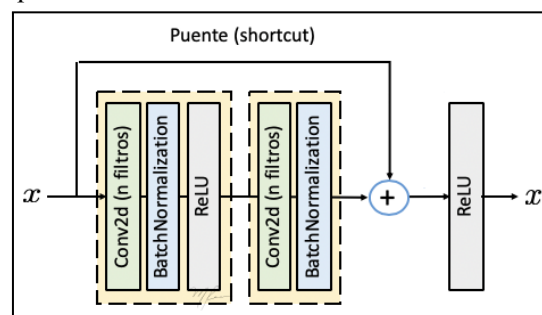
- **Train:** Alrededor de 14,000 imágenes para entrenamiento.
- **Test:** Alrededor de 3,000 imágenes para evaluación del modelo.
- **Prediction:** Alrededor de 7,000 imágenes para predicción final.

Objetivo:

Utilizando el dataset proporcionado, el objetivo es construir y comparar el rendimiento de distintos modelos de clasificación de imágenes utilizando redes neuronales convolucionales y densas que puedan clasificar con precisión las imágenes de escenas naturales en una de las seis categorías mencionadas anteriormente.

Los modelos a diseñar son:

- Modelo con capas densas.
- Modelo con capas convolucionales y densas.
- Modelo que incluya bloques residuales identidad:



Bloque residual identidad

- Modelo que utilice como backbone alguna de las arquitecturas disponibles en TensorFlow (transfer learning): https://www.tensorflow.org/api_docs/python/tf/keras/applications

Se debe entrenar y evaluar cada modelo utilizando técnicas adecuadas de validación y métricas de evaluación de clasificación.

Entrega:

La entrega debe incluir:

Código fuente de la solución implementada en Google Colab, que incluya:

- Análisis previo y preprocesamiento del set de datos.
- Definición y entrenamiento del modelo.
- Resultados de la evaluación del modelo, incluyendo métricas de desempeño y visualizaciones relevantes.

Nota: el código debe estar debidamente documentado con comentarios explicativos para que el trabajo sea fácilmente comprensible para otros revisores.