



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 médicos de distintos pacientes. El objetivo es determinar si el paciente padece o no de diabetes en base a ciertos indicadores médicos. Todos los pacientes del dataset son de sexo femenino.

Dataset:

El dataset proporcionado incluye las siguientes variables para cada paciente:

- **Embarazos:** Número de veces embarazada
- **Glucosa:** Concentración de glucosa en plasma a las 2 horas en una prueba de tolerancia oral a la glucosa
- **Presión arterial:** Presión arterial diastólica (mm Hg)
- **Espesor de la piel:** Espesor del pliegue cutáneo del tríceps (mm)
- **Insulina:** Insulina sérica a las 2 horas (μ U/ml)
- **IMC:** Índice de masa corporal ($\text{peso en kg}/(\text{altura en m})^2$)
- **Función de pedigree de la diabetes:** Valor que evalúa la probabilidad de diabetes basada en antecedentes familiares.
- **Edad:** Edad (años)
- **Resultado:** variable objetivo (0 para diabetes negativo, 1 para positivo).

Objetivo:

Utilizando el dataset construido, el objetivo es construir un modelo de clasificación utilizando redes neuronales que pueda predecir con precisión si el paciente en cuestión posee o no diabetes. Se debe entrenar y evaluar el 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.

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:

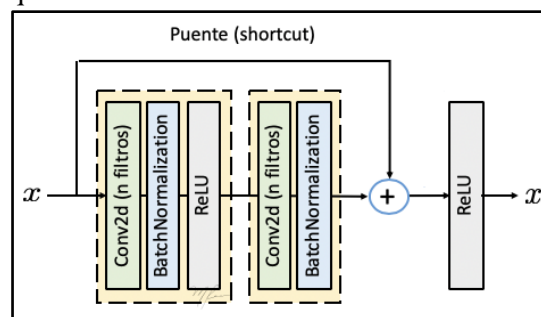
- **Train:** Alrededor de 14,000 imágenes para entrenamiento.
- **Test:** Alrededor de 3,000 imágenes para evaluación del modelo.
- **Prediction:** Alrededor de 7,000 imágenes para predicción final.

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.