**Examen Consultor de Machine Learning**

**Instrucciones**

* Envía tus respuestas a más tardar 12 horas después de recibir el correo
* Las respuestas podrán ser enviadas en un PDF, un notebook, o con un link a tu repositorio de GitHub
* Debes especificar y comentar de manera clara el código y los incisos respondidos

**Sección A. Teoría**

**Redes Neuronales Convolucionales**

1. ¿Como medir el nivel de confianza en un modelo?
2. ¿Cuál es la utilidad de las técnicas: ¿Transfer learning y Fine Tuning?
3. ¿Para qué sirve la capa de MaxPooling?
4. ¿Qué parámetros requiere como mínimo la capa de Dropout en Keras?
5. ¿Cuáles son los parámetros que aprender la red neuronal en las capas convolucionales?
6. Regresión, Clasificación Binaria o Clasificación Multiclase ¿cuál es el problema más difícil de aprender para una red neuronal? Justifica tu respuesta
7. ¿Cuál de los siguientes métodos NO evita que un modelo se sobreajuste al conjunto de entrenamiento? Explica tu respuesta

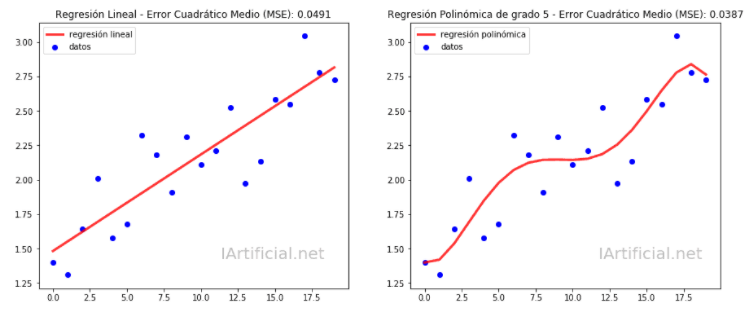
Early Stooping Dropout Data Augmentation Pooling

1. ¿Cuál de las siguientes técnicas realiza una operación similar al dropout? Explica tu respuesta

Bagging Boosting Stacking Ninguna

**Intuición Gráfica**

1. Analicemos las siguientes imágenes correspondientes a la solución de un problema de regresión



Ambas son soluciones a un problema de regresión, el de la izquierda es un modelo de regresión lineal. Tiene un error cuadrático medio de 0.0501, el de la derecha es un modelo de regresión polinómica de grado 5. Tiene un error Cuadrático medio de 0.0289

¿Cuál de los dos modelos tendrá mejores resultados ante su exposición a nuevos datos? Justifica tu respuesta

1. Tenemos una red neuronal convolucional, entrenada para reconocer los caracteres H y T como se muestra a continuación:

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

¿Cuál podría ser la salida esperada por la red con el input de la tercera cuadrilla? Justifica tu respuesta

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

**Sección B. Practica Machine Learning**

**Ejercicio 1**

**Objetivo:** Construir un sistema muy sencillo para la clasificación de Rostros.

**Instrucciones:** Contamos con un conjunto de datos link que contiene datos que contiene dos clases: Clase Positiva: Imágenes de Rostro, Clase Negativa: Imágenes Variadas. Cada clase cuenta con 115 imágenes totales. Cada imagen tiene un tamaño de 24x24 pixeles, pero están almacenadas en vectores de 1x576.

1. Calcule los componentes principales de la BD
2. Calcule la media de la base de datos, que es el rostro promedio y muéstrelo
3. Normalice la base de datos a una media=0
4. Reduzca la base de datos a 15 dimensiones
5. Proponga un algoritmo de Clasificación y sus métricas de evaluación

**Ejercicio 2.**

**Objetivo:** Describir y limpiar el dataset otorgado, después proponer un sistema de clasificación al problema.

**Instrucciones:** Contamos con el dataset “DataCleanClassificationBoth.csv” con historiales crediticios de los clientes del banco N, con el fin de realizar la predicción para autorizar o negar créditos.

1. Genera un diagrama de bloques para predecir la variable “Response” con la mayor exactitud posible.
2. Realizar la preparación de datos: carga, limpieza y procesamiento.
3. ¿Cuáles fueron los pasos aplicados en el data cleaning?
4. Proponer uno o varios algoritmos de clasificación, y como deberían ser probados y comparados
5. Elige el mejor modelo y justifica tu respuesta
   1. Si utilizas pasos adicionales de optimización para el modelo, asegúrate de incluirlos

**Sección C. Práctica Deep Learning**

**Objetivo:** Implementar un clasificador de razas de perros utilizando Deep Learning

**Instrucciones:** En el archivo standford\_db.zip se encuentra una muestra de la base de datos Standford Dogs Dataset, el cual está conformado por imágenes de 15 clases distintas de perros, con el fin de poder categorizar las imágenes de manera correcta.

1. Genera un diagrama, de descripción a modo de bloques, con las etapas necesarias para cumplir el objetivo del ejercicio
2. ¿Cuál es la distribución de la población de la base de datos?
3. ¿La base de datos es balanceada o desbalanceada? ¿Qué técnica se aplicaría para resolver este problema?
4. ¿Es necesario emplear técnicas de aumento de datos? Si/No, por qué?
5. ¿Qué arquitectura de red elegirías para el clasificador? Sustenta tu respuesta