

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ciencias y Sistemas
Inteligencia Artificial 1
Vacaciones de diciembre de 2020
Catedrático: Ing. Jorge Gutiérrez
Auxiliar: Nery Galvez



Práctica 2

Tabla de Contenido

OBJETIVOS	2
OBJETIVO GENERAL.....	2
OBJETIVOS ESPECÍFICOS	2
DESCRIPCIÓN DE LA SOLUCIÓN.....	2
CARGAR IMÁGENES.....	3
ANALIZAR IMÁGENES.....	4
ALGORITMO.....	4
CONJUNTOS DE DATOS.....	4
HÍPER PARÁMETROS	5
ENTRENAMIENTO	5
ENTRENAMIENTO DE MODELOS	5
SELECCIÓN DEL MODELO	6
ACLARACIONES	7
CONSIDERACIONES	7
FORMA DE ENTREGA	7
ENTREGABLES	7
FECHA DE ENTREGA.....	7

Objetivos

Objetivo General

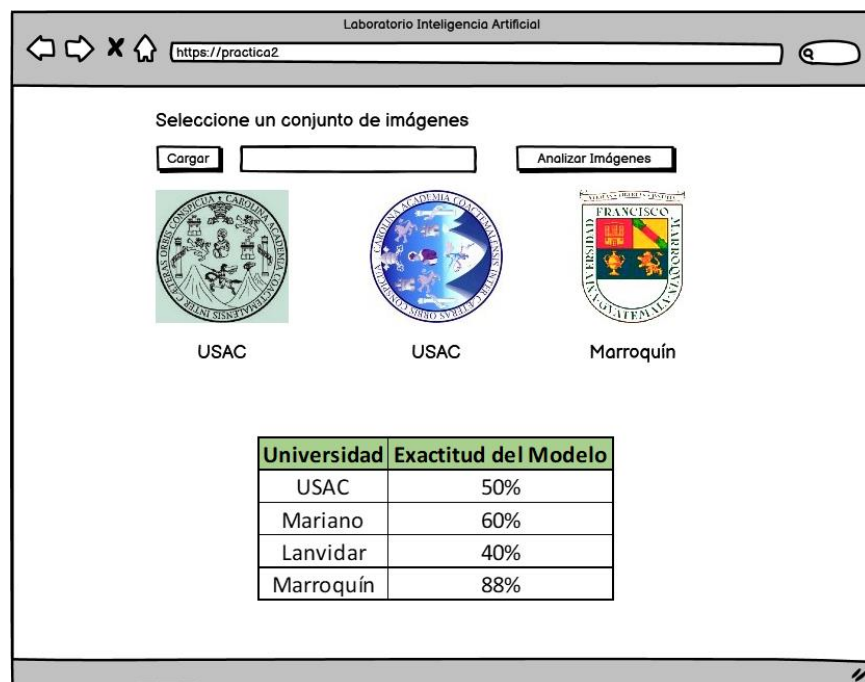
Implementar una aplicación web donde se apliquen los conceptos adquiridos en el laboratorio de Inteligencia Artificial 1.

Objetivos Específicos

- Implementar un algoritmo de Machine Learning que permita realizar predicciones sobre un conjunto de datos.
- Resolver un problema de clasificación implementando un algoritmo de regresión logística.

Descripción de la Solución

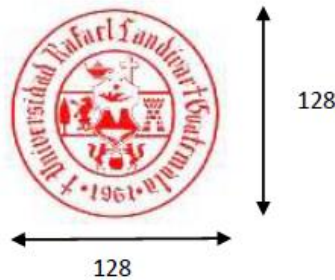
El estudiante debe de crear una aplicación web que utilizando un modelo de aprendizaje automático reciba un conjunto de imágenes y tenga la capacidad de decir a qué universidad pertenece el escudo. A continuación, se muestra una posible interfaz y se definen las funcionalidades de la aplicación.



Cargar Imágenes

La aplicación permitirá cargar desde una hasta N imágenes para poder ser analizadas por el modelo y que este pueda predecir a que universidad corresponde cada imagen.

Las imágenes deberán cumplir con el siguiente estándar:



Cada imagen será a color y tendrá un tamaño de 128x128. El nombre de la imagen tendrá el siguiente formato:

<Universidad>_<número imagen>.jpg

Por ejemplo, supongamos que cargamos un conjunto de 5 imágenes. Los nombres de las imágenes que cargamos podrían ser:

USAC_1.jpg

Mariano_2.jpg

USAC_3.jpg

Landivar_4.jpg

Marroquin_5.jpg

Con el formato anterior nos aseguramos que las imágenes de entrada lleven la respuesta correcta en el nombre y por medio de los números no pueden existir imágenes con el mismo nombre.

Analizar Imágenes

Esta función mostrará los resultados devueltos por el modelo, el resultado a mostrar es el siguiente:

- Si se cargaron de 1 a 5 imágenes, en esta sección se mostrarán las imágenes cargadas junto con la predicción del modelo de a qué institución pertenecen.
- Si se cargaron de 6 a N imágenes, lo que se mostrara es una tabla con los resultados generales (porcentaje de imágenes de cada institución en las que acertó el modelo con la predicción). Para sacar el porcentaje tienen que comparar la respuesta correcta, que viene en el nombre la imagen, contra la respuesta que su algoritmo está dando, esto lo hacen para cada imagen y sacan el porcentaje de aciertos de su modelo.

Algoritmo

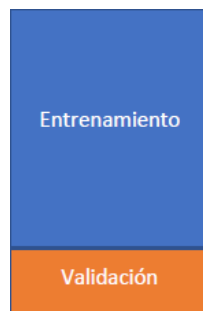
El modelo que utilizará la página web será entrenado por medio de el algoritmo de la regresión logística. Es decir, las imágenes que se cargaran en la página no se van a utilizar para entrenar el modelo, el modelo ya debe estar entrenado con el siguiente conjunto de imágenes:

<https://drive.google.com/drive/folders/1lfAApJ2rjhUd6sPi67SJZQxp9v0kS39U?usp=sharing>

.

Conjuntos de datos

Van a tomar el conjunto de datos anterior y lo van a dividir en dos secciones, que son las siguientes:



La cantidad de datos que se utilizará para el entrenamiento y la cantidad para la validación quedará a discreción del estudiante.

Híper parámetros

Los hiper parámetros que se utilizaran para configurar el entrenamiento del algoritmo son los siguientes:

- Tasa de aprendizaje
- Tasa de regularización
- Cantidad de iteraciones

Entrenamiento

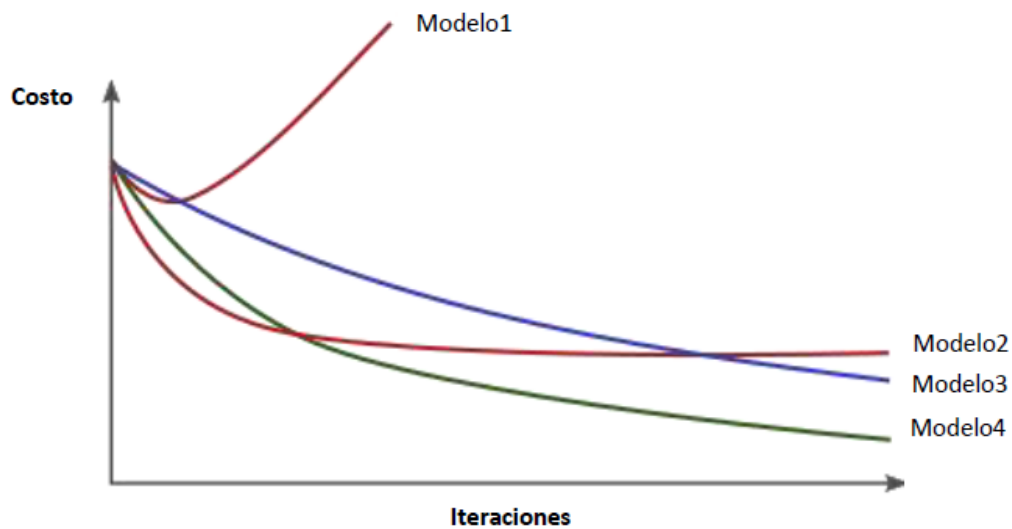
Entrenamiento de modelos

Van a crear cinco modelos para cada universidad, cada uno con una configuración distinta de hiper parámetros. Por ejemplo, un modelo podría tener la siguiente configuración:

- Tasa de aprendizaje = 0.01
- Tasa de regularización = 0.5
- Cantidad de iteraciones = 1500
-

Van a entrenar los 5 modelos con el conjunto de entrenamiento. Y graficarán cómo fue el comportamiento de los 5 modelos en un mismo plano. Es decir, como fue cambiando el valor del costo conforme iban pasando las iteraciones.

Por lo tanto, tendrán una gráfica con 5 modelos por cada universidad.



Gráfica de los modelos entrenados para la USAC

Selección del modelo

Después de realizar los modelos de las universidades evaluarán el rendimiento de los modelos en el conjunto de entrenamiento y en el conjunto de validación. Para ello utilizarán la siguiente formula:

$$Exactitud = \frac{Cantidad\ de\ aciertos}{Cantidad\ de\ datos\ en\ el\ conjunto} * 100$$

Los resultados los guardarán en un archivo de texto. El formato del archivo queda a discreción del estudiante. Un ejemplo del archivo de entrada podría ser:

```
Modelo1
Entrenamiento = 80%
Validación = 40%
...
Modelo5
Entrenamiento = 70%
Validación = 50%
```

En base a los resultados que observen en las gráficas y en el rendimiento de cada modelo elegirán el modelo que crean que sea mejor y ese es el que utilizara el prototipo (la página web).

Aclaraciones

- Para la calificación ya tendrían que haber entrenado los 5 modelos y haber generado la gráfica del entrenamiento y el archivo con los resultados de los cinco modelos. Así como haber elegido ya a su mejor modelo para utilizarlo en la página web.
- Durante la calificación solamente se les pedirá entrenar un modelo más para validar la implementación del algoritmo. Los valores de los hiper parámetros serán dados en la calificación.
- Se hace énfasis en la selección del mejor modelo ya que en la hoja de calificación habrá una sección que calificará el rendimiento del mejor modelo presentado.

Consideraciones

- La aplicación deberá desarrollarse utilizando Python 3.
- El framework para la aplicación web queda a discreción del estudiante.
- El proyecto se realizará de manera individual, si se detecta algún tipo de copia el laboratorio quedará anulado.
- Se debe de crear un repositorio para llevar control de su proyecto.

Forma de Entrega

Entregables

- El entregable será el enlace al repositorio del proyecto, recordar que el auxiliar debe de tener siempre acceso a dicho repositorio, de lo contrario se anulará la nota obtenida en el proyecto.

Fecha de entrega

Miércoles 23 de diciembre de 2020, hasta las 23:59