Project: Plant Seedlings Image Classification using CNN

Learning Objectives

- To implement the techniques learnt as a part of the course.
- Pre-processing of image data.
- Visualization of images.
- Building CNN
- Evaluate the Model

Data Description

greatlearning

- You are provided with a training set and a test set of images of plant seedlings at various stages of grown.
- Each image has a filename that is its unique id.
- The dataset comprises 12 plant species.
- The goal of the competition is to create a classifier capable of determining a plant's species from a photo.
- **Dataset :** Link to the Kaggle project site: https://www.kaggle.com/c/plant-seedlings-classification/data

Steps to follow

greatlearning

- Import the libraries, load dataset, print shape of data, visualize the images in dataset.
- Data Pre-processing:
  - Normalization.
  - Gaussian Blurring.
  - Visualize data after pre-processing.
- Make data compatible:
  - Split the dataset into training, testing, and validation set.

Steps to follow

greatlearning

- Make data compatible:
  - Split the dataset into training, testing, and validation set.
    - (Hint: First split train images and train labels into training and testing set with test_size = 0.3. Then further split test data into test and validation set with test_size = 0.5)
    - [Read the note at the end of the problem statement for the reason behind using the train images for using for training and testing set.]
  - Reshape data into shapes compatible with Keras models.
  - Convert labels from digits to one hot vectors.
  - Print the label for y_train[0].

Steps to follow

greatlearning

- Building CNN:
  - Define layers.
  - Set optimizer and loss function. (Use Adam optimizer and categorical crossentropy)
- Fit and evaluate model and print confusion matrix.
- Visualize predictions for x_test[2], x_test[3], x_test[33], x_test[36], x_test[59].

Note:

- Download the train images from the Kaggle dataset, don’t download the test images.
- As the test dataset is not labeled, so you won’t be able to calculate the accuracy score.
- So use train images and train labels only to split further into training and testing set during your model building.

greatlearning

HappyLearning

Questions?