Schneider Electric Challenge

Sravan (vskvarma1305@gmail.com)

Data Exploration

- Three types of Images
  - Plantation
  - Grassland / Shrubland
  - Smallholder Agriculture
- Labels are not ordered by Image numbers and so each image is loaded by path in CSV file and assigned respective label.
- Data is unbalanced among different classes. Grassland / Shrubland has just ~ 10%.

Data Preparation

- Images are augmented by rotating and flipping. It gives more variation of data to train the model the better.
  - Rotated 90 degrees clockwise and counterclockwise
  - Flipped vertically and horizontally
- Train data is further split (80:20) into train and validation
  - Train data will be used to train while validation will be used to estimate the performance of model
- Images are scaled to improve learning rate of model

Model

- VGG16 architecture is used and trained on *imagenet* dataset. Features from this model are extracted for Transfer Learning.
- PCA is applied on features to see the clusters from Transfer Learning. There is a large overlap from the feature set obtained from Transfer Learning indicating the need to improve the model.
- We apply a ReLU(Rectified Linear Unit) and Softmax regression on the features of Transfer Learning to learn from our Train Dataset.
- F1 Score of ~0.60 shows that there is room for improvement and We use the latitude, longitude and year features and apply a Random Forest Classification.