


A decorative graphic on the left side of the slide consists of two overlapping parallelograms. The front one is blue and the back one is a light green. They are positioned diagonally, with the blue one partially covering the green one.

Hall B: Week 2

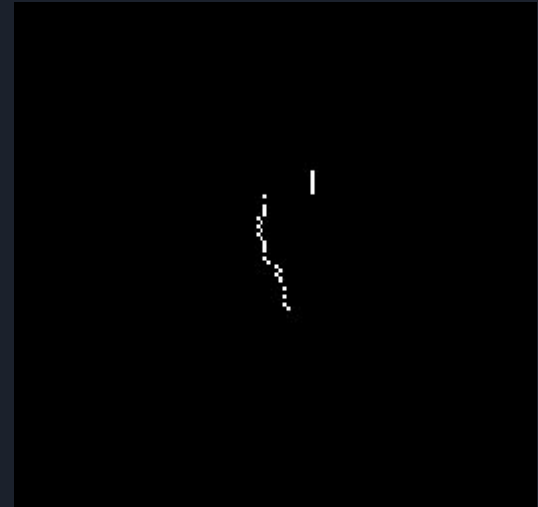
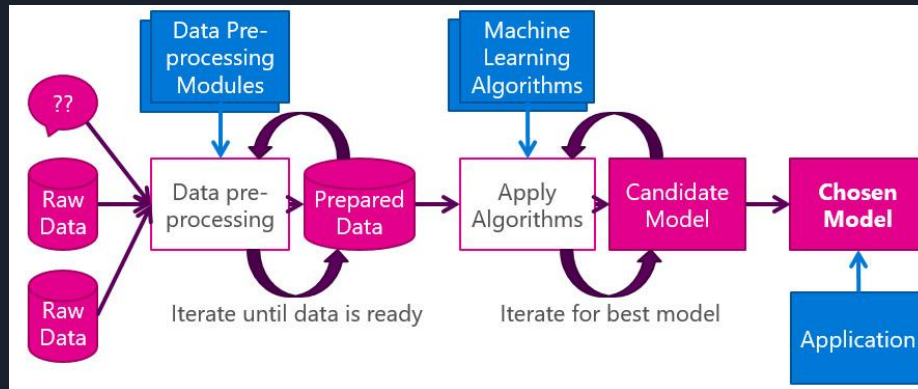
Jose Cruz and Andrew Hoyle

Goals for the past week

- 
- Process the data to run it through a pre-trained model
 - Find a pre-trained model to run our data through
 - Run the data generated by the model through a logistic regression algorithm

Data Processing

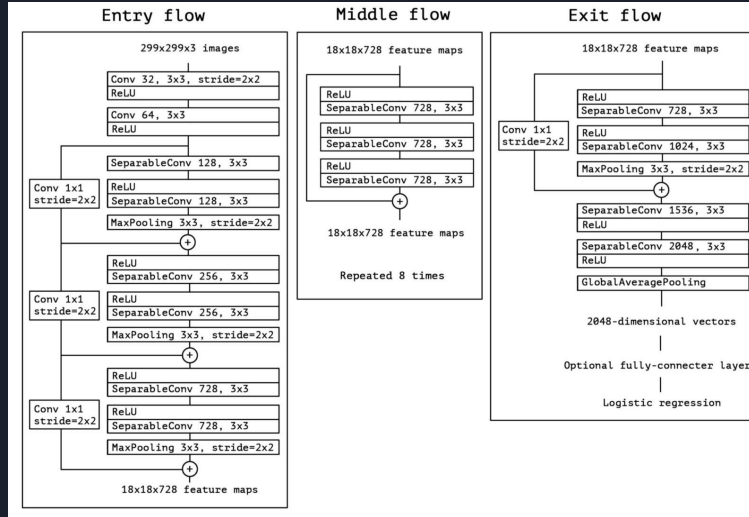
- resized images (112,112,3) and converted pixel data to numpy arrays
- shuffled data and separated into testing and training sets. (roughly 75/25 split)



112X112 image

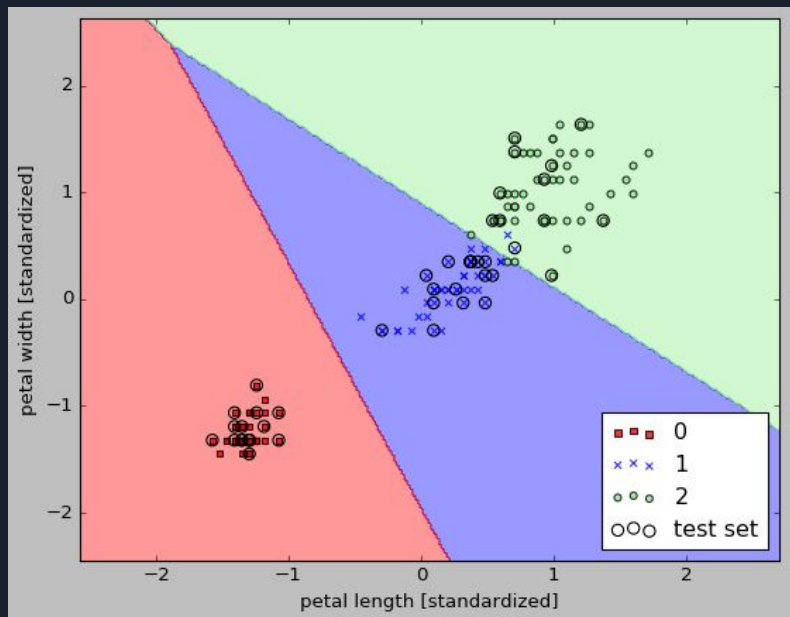
Pre-Trained Model

- used the pre-trained conv net Xception to generate vectors from data
 - 'froze' the top layer to prevent the model from making its own predictions.
 - Converted our data to a batched Tensorflow dataset (easier for the computer to handle)
 - Generated a numpy vector for each of the images



Logistic Regression

- Used native scikit-learn logistic regression algorithms to make predictions from data vectors



Next Steps

- Fix batching problems
- Regularize our model to maximize prediction accuracy and avoid over/underfitting
- Optimize calculation speed (switch from CPU to GPU)

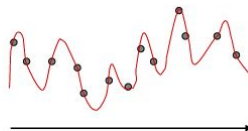
Regularization

- The minimization

$$\min_f |Y_i - f(X_i)|^2$$

may be attained with zero errors.

But the function may not be unique.



- Regularization

$$\min_{f \in H} \sum_{i=1}^n |Y_i - f(X_i)|^2 + \lambda \|f\|_H^2$$

- Regularization with smoothness penalty is preferred for uniqueness and smoothness.
- Link with some RKHS norm and smoothness is discussed in Sec. IV.

