Machine Learning for Computer Vision:
Coursework 1 - Randomised Decision Forests

Tormento, Marion
Imperial College London
marion.tormento17@imperial.ac.uk

McLaughlin, Edward
Imperial College London
edward.mclaughlin17@imperial.ac.uk

Abstract

In this paper, the properties and power of Random Forests (RFs) is explored. In question one and two classifier is implemented on a 3-class spiral data set, and the resultant tree is tested on sample points and a blanket array of points. In this part, the entirety of the code (except for the getData.m file) was created by the authors and can be supplied at request. In question three, RFs are use to create a codebook with the purpose of image categorisation of the CalTech101 images provided. Firstly, a kmeans codebook was built, subsequently its performance was compared with an RF codebook, in each case using an RF classifier.

Introduction

Question 1

Figure 1 shows the result from bagging the data into four bags *with* replacement. The size of the bags is determined by equation 1, which results in the bags being roughly 63.2% of the original data set i.e. the bootstrap sampling fraction (BSF) ($\approx$ 63.2%).

$$BSF = 1 - \frac{1}{e} \tag{1}$$

It can be seen that while the bags display similar, but not identical, histograms. Thus a source of randomness is introduced into the algorithm in this initial step. Each bag is used as the root node for each tree in the Randomised Forest (RF). The root node is split in an optimal manner, explained later in the section, to maximise the Information Gain produced by the split function. The IG calculation is given in equation 2.

$$I(S,\Theta) = H(S) - \sum_{i \subset L,R} \frac{|S^i|}{|S|} H(S^i) \tag{2}$$

where, S is the total set of data points, $S\hat{}L$ and $S\hat{}R$ are the two children sets of data defined by the split function in question and H is the Shannon entropy defined by

$$H(S) = -\sum_{c \subset C} p(c) log(p(c)) \tag{3}$$

where S is the set of data points, c denotes the class in question and C denotes the full set of classes considered.

From this split, two children are produced. Each child, if not deemed to be a leaf node, will also be optimally split. This process continues until all branches end with a leaf node. A leaf node is created if: a) there are fewer than 10 points in the child, b) there is only one class left in the child, c) the child node is part of the final level of the tree.

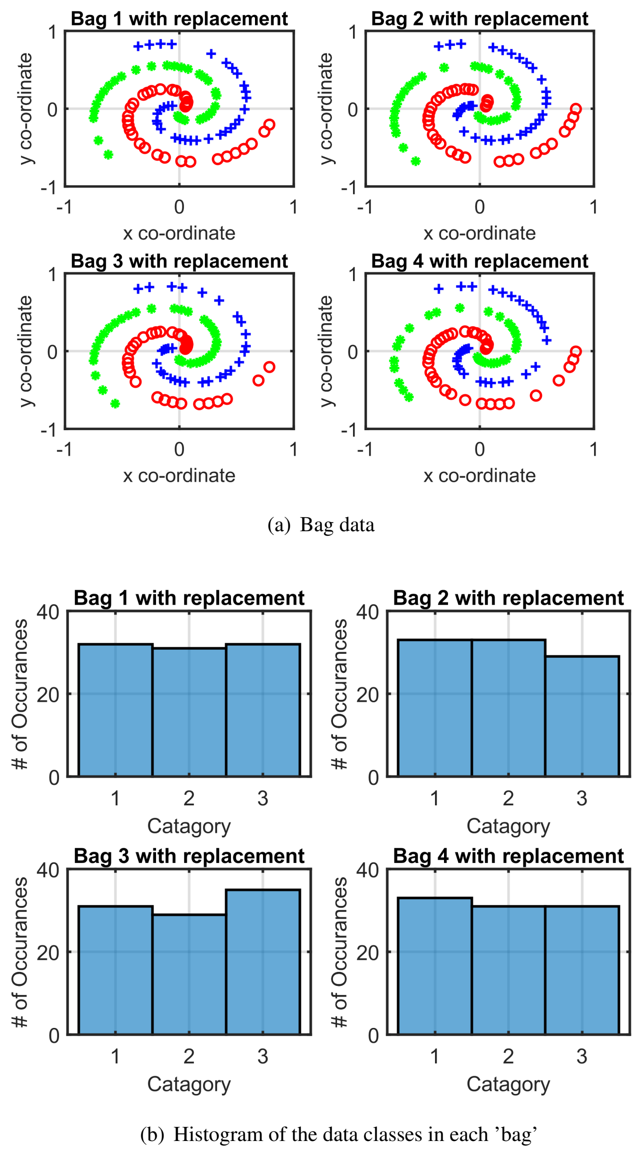


Figure 1. Bootstrap Aggregating (Bagging) with replacement

The process of finding an optimal split node is shown graphically in figure 2 and is performed as follows: 1) parameter ρ is defined. 2) each split node function type (either x- or y-axis aligned or linear) is implemented with random parametrisations a set number of times, dependent on ρ. 3) The split function calculated to have the highest Information Gain (IG) is selected as the optimal split function. This manner of selecting optimal split functions based on random parametrisation introduces a second course of randomness into the system.

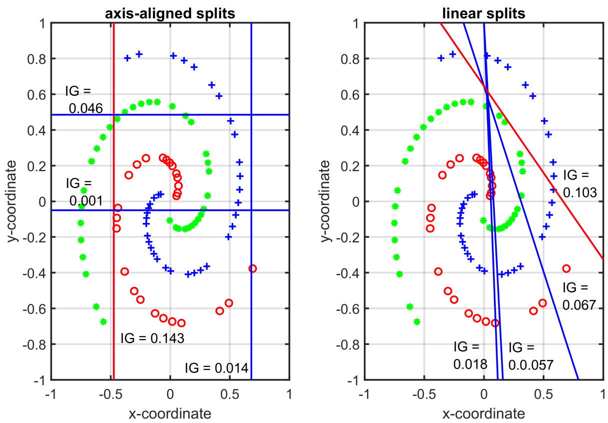


Figure 2. Example split functions shown (ρ = 2): On the left the, 2 functions in both x- and y-axis aligned, on the right, 4 functions in linear aligned. Next to each line is its Information Gain (IG), the function with the highest IG in each case is highlighted displayed in red. These two functions are then compared and the best is selected for the split node.

Question 2

To evaluate the performance of the RF for the chosen parameters, first several points are taken and their class probabilities are visualised using histograms (figures 1 - 1), then a dense 2D grid of points is used to visualise a colour data space. To this end, figure 3 shows a map of the classification at each point, whereas figure 4 visualises the probability distribution of point by weighting their RBG values by the probability that they belong to the red, green or blue class respectively.

The parameters found to give acceptable results in a fast training time were... .

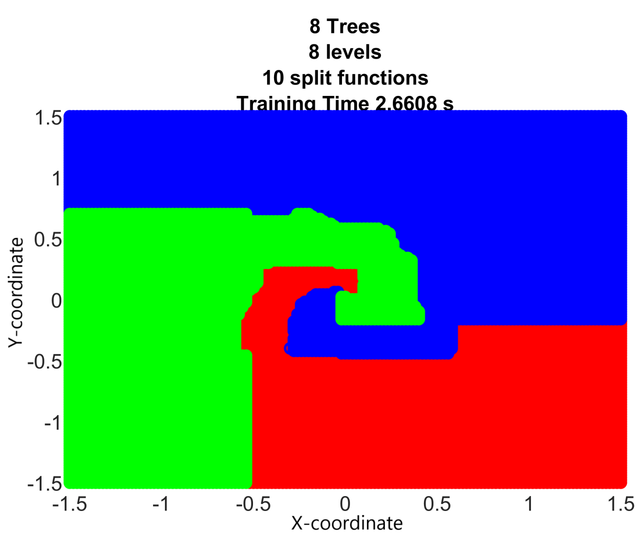


Figure 3.

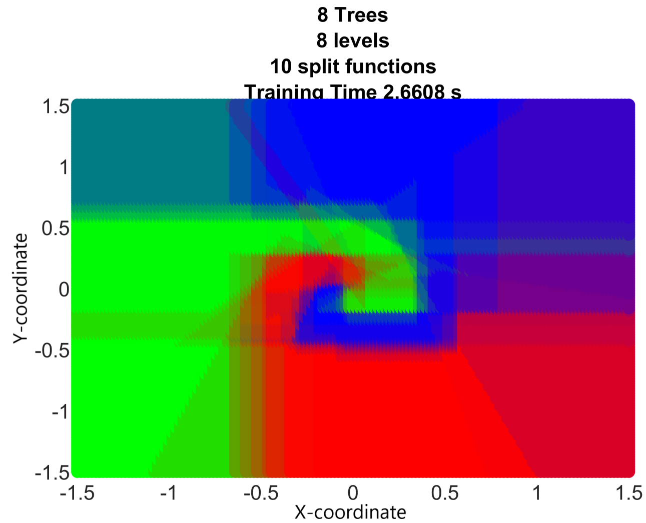


Figure 4.

Question 3

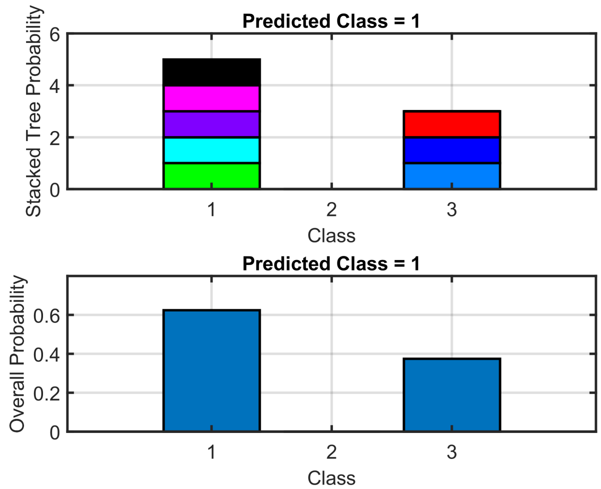
Part 1

Part 2

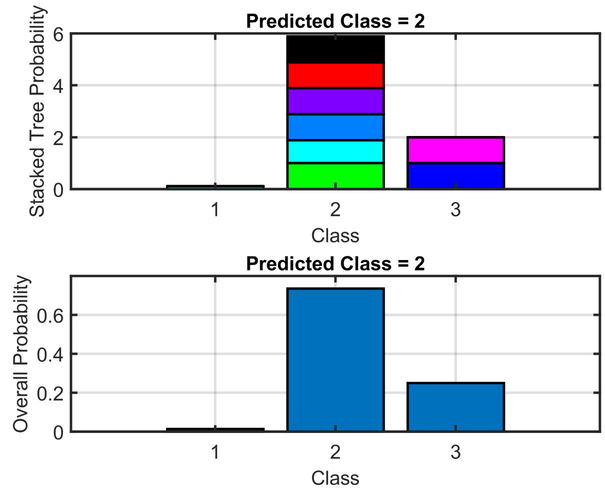
Part 3

Optional Further Discussion

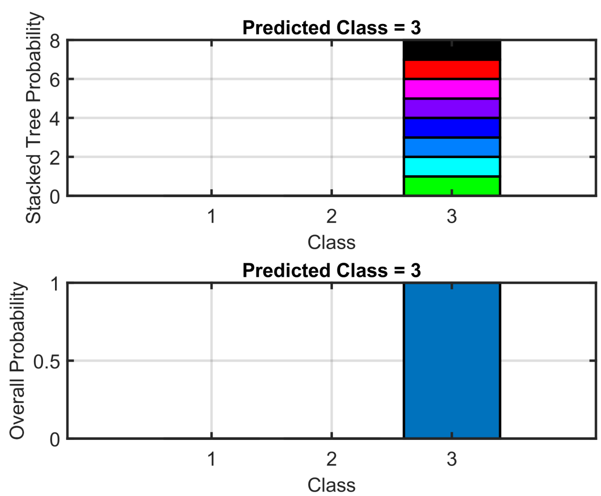
1 Appendix



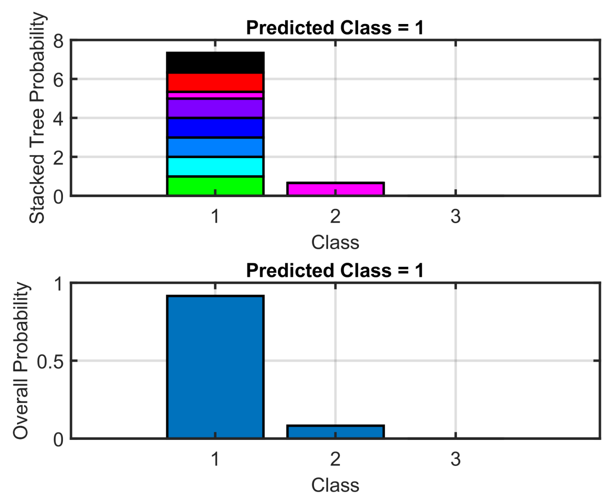
]
(a) Point A: [-0.7, -0.5



]
(b) Point B: [0.4, 0.3



]
(c) Point C: [-0.7, 0.4



]
(d) Point D: [0.5, -0.5

Figure 5. Visualisation of the class probability of four sample points.