Pattern and Speech Recognition WS2015-16
Exercise 8

Atanas Poibrenski(2554135), Marimuthu Kalimuthu(2557695), Furkat Kochkarov(2557017)

January 9, 2016

Decision Trees

1 Loading the data

Ex.1 Done.

Ex.2 See "data_preparation.m"

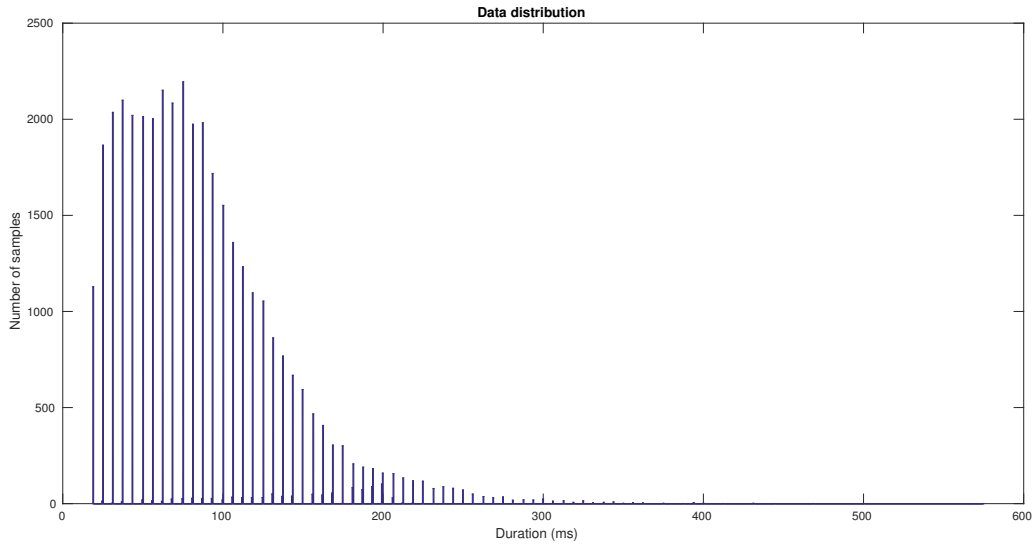

Figure 1: Histogram plot of phoneme durations

Ex.3 **Bonus**:
Yes, we can use Gaussian Mixture Models for duration modeling. We create 60 components which is the total number of phoneme types(i.e., Vowel, Consonant) and fit a Gaussian to each. Then a new phoneme comes we predict the cluster with highest probability which contains the phoneme.

Ex.4 **Bonus**:
In GMM, we always have a certain probability of a phoneme belonging to particular cluster. In decision trees, a phoneme either belongs to certain type or not.(i.e. it is a binary decision).

Ex.5 Divide dataset:
We use matlab's 'dividerand' function which gives random indices. (i.e. shuffle the data). Then we get the following counts:

> Training set (90%) – 34980 samples
> Validation set (5%) – 1943 samples
> Test set (5%) – 1943 samples

2 Decision Tree clustering

2.1 Preliminary work

Ex.6 data structure:
We use matlab 'struct' which has three attributes 'value', 'left', and 'right'. Since we will be using recursion for creation of the tree, it is convenient to use it.

2.2 Criterion

Ex.7 find_question:
See 'find_question.m'. We calculate the entropy impurity for each question and choose the one with minimum impurity.

Ex.8 eval_tree:
See 'eval_tree.m'. It is calling 'classify_by_tree' for each data point in the given corpus. And it is calculating the RMSE.

2.3 Training the tree

Ex.9 'train_tree' See 'train_tree.m'. It is creating the tree in a recursive manner with a fixed depth for the tree.
In 'training.m', we call 'train_tree.m' in an iterative deepening manner. This way we get trees with different depths, so that we can apply the stopping criteria where we evaluate two trees of different depths.

3 Some analysis

Ex.10 The same as we did in Ex.5 will work here too.

Ex.11 Our training of the tree is not working as desired.

Ex.12 Our training of the tree is not working as desired.