Voice Recognition through Machine Learing

Raktim Gautam Goswami[1], Abhishek Bairagi[2] & G V V Sharma[3]

Contents

***Abstract*—This manual shows how to develop a voice recognition algorithm and use it to control a toycar.**

1 Dataset

1.1 Draw the block diagram of the AI-ML system for the toycar.
Solution: See Fig. 1.1

1.2 Record 'forward' 80 times using you phone and save as 'forwardi.wav' for $i = 1, \ldots, 80$. The recording duration should be between 1-3 seconds.

1.3 Repeat by recording 'left', 'right', 'back' and 'stop'. Make sure that the audio files for each command are in separate directories. Download the following directory for reference

```
svn checkout https://github.com/gadepall/
    EE1390/trunk/AI−ML/audio_dataset
```

1.4 Use the following script to generate a dataset for 'back' command. Explain through a block diagram.

```
https://raw.githubusercontent.com/gadepall/
    EE1390/master/AI−ML/codes/250files.py
```

Solution: The datasets are generated through zero padding. The diagram in Fig. 1.4 explains how this is done for the back command.

The authors are with the Department of Electrical Engineering, Indian Institute of Technology, Hyderabad 502285 India . e-mail: 1. ee17btech11004@iith.ac.in, 2. ee17btech11051@iith.ac.in, 3. gadepall@iith.ac.in

1.5 Suitably modify the above script to generate similar datasets for 'left', 'right', 'stop' and 'forward'.

1.6 Summarize the datasets generated through a table.
Solution: See Table 1.6

2 Linear Regression: Least Squares

2.1 Draw the block diagram for the ML algorithm
Solution: See Fig. 2.1

2.2 List the reference vectors for all the voice commands.
Solution: See Table 2.2.

2.3 The sigmoid function is defined as

$$s(x) = \frac{1}{1 + e^{-x}} \tag{2.1}$$

Sketch $s(x)$.
Solution: The following code plots $s(x)$ in Fig. 2.3.

2.4 Show that $0 < s(x) < 1$.
Solution: $s(x)$ is useful for transforming large values to a value between 0 and 1.

2.5 Formulate a regression model for the voice recognition system.
Solution: Let $\mathbf{x}$ be the voice command. The model used is

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b} \tag{2.2}$$

where $\mathbf{W}$ is the weight matrix and $\mathbf{b}$ is the bias vector. Ideally, the output $\mathbf{y}$ should be one of the reference vectors in Table 2.2.

2.6 Frame an optimization problem for estimating $\mathbf{W}$ and $\mathbf{b}$.
Solution: This is done by considering the cost function

$$\min_{\mathbf{W},\mathbf{b}} J\left(\mathbf{W}, \mathbf{b}\right) = \frac{1}{2}\|\mathbf{y} - \hat{\mathbf{y}}\|^2 \tag{2.3}$$

Commands	Input	Output / Input file	Conditioned	Training	Testing
Back	80	250	20000	16000	4000
Forward	80	250	20000	16000	4000
Left	80	250	20000	16000	4000
Right	80	250	20000	16000	4000
Stop	80	250	20000	16000	4000
	Total		**100000**		

TABLE 1.6: File calculus

Command	Reference vector
Forward	$\begin{pmatrix}1 & 0 & 0 & 0 & 0\end{pmatrix}^T$
Back	$\begin{pmatrix}0 & 1 & 0 & 0 & 0\end{pmatrix}^T$
Left	$\begin{pmatrix}0 & 0 & 1 & 0 & 0\end{pmatrix}^T$
Right	$\begin{pmatrix}0 & 0 & 0 & 1 & 0\end{pmatrix}^T$
Stop	$\begin{pmatrix}0 & 0 & 0 & 0 & 1\end{pmatrix}^T$

TABLE 2.2: Reference vectors

where

$$\hat{\mathbf{y}} = \mathbf{W}\mathbf{x} + \mathbf{b} \tag{2.4}$$

3 Gradient Descent

3.1 Let $\mathbf{W}$ be 2×2 and $\mathbf{x}, \mathbf{y}$ be 2×1. Show that

$$\frac{\partial \left(\mathbf{y}^T \mathbf{W}\mathbf{x}\right)}{\partial \mathbf{W}} = \mathbf{y}\mathbf{x}^T \tag{3.1}$$

3.2 Given that $\mathbf{b}$ is 2×1, show that

$$\frac{\partial \|\mathbf{W}\mathbf{x} + \mathbf{b}\|^2}{\partial \mathbf{W}} = 2\left(\mathbf{W}\mathbf{x} + \mathbf{b}\right)\mathbf{x}^T \tag{3.2}$$

3.3 Find

$$\frac{\partial \|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\partial \mathbf{W}} \tag{3.3}$$

Solution:

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \|y\|^2 - 2\mathbf{y}^T\left(\mathbf{W}\mathbf{x} + \mathbf{b}\right) + \|\mathbf{W}\mathbf{x} + \mathbf{b}\|^2 \tag{3.4}$$

From (3.1) and (3.2)

$$\frac{\partial \|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\partial \mathbf{W}} = -2\left(\mathbf{y} - \hat{\mathbf{y}}\right)\mathbf{x}^T \tag{3.5}$$

3.4 Show that

$$\frac{\partial \mathbf{y}^T \mathbf{b}}{\partial \mathbf{b}} = \mathbf{y} \tag{3.6}$$

3.5 Show that

$$\frac{\partial \|\mathbf{b}\|^2}{\partial \mathbf{b}} = 2\mathbf{b} \tag{3.7}$$

3.6 Show that

$$\frac{\partial \|\mathbf{W}\mathbf{x} + \mathbf{b}\|^2}{\partial \mathbf{b}} = 2\left(\mathbf{W}\mathbf{x} + \mathbf{b}\right) \tag{3.8}$$

3.7 Show that

$$\frac{\partial \|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\partial \mathbf{b}} = -2\left(\mathbf{y} - \hat{\mathbf{y}}\right) \tag{3.9}$$

3.8 $\mathbf{W}$ and $\mathbf{b}$ can be estimated from (2.3) using

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \frac{\alpha}{2}\frac{\partial J\left(\mathbf{W}, \mathbf{b}\right)}{\partial \mathbf{W}} \tag{3.10}$$

$$\mathbf{b}(n+1) = \mathbf{b}(n) - \frac{\alpha}{2}\frac{\partial J\left(\mathbf{W}, \mathbf{b}\right)}{\partial \mathbf{b}} \tag{3.11}$$

Show that (3.10) can be expressed as

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \alpha\left(\mathbf{y} - \hat{\mathbf{y}}\right)\mathbf{x}^T \tag{3.12}$$

$$\mathbf{b}(n+1) = \mathbf{b}(n) + \alpha\left(\mathbf{y} - \hat{\mathbf{y}}\right) \tag{3.13}$$

3.9 Show that

$$s'(x) = (1 - s(x))\, s(x) \tag{3.14}$$

3.10 Suppose

$$J\left(\mathbf{W}, \mathbf{b}\right) = \frac{1}{2}\|\mathbf{y} - s\left(\hat{\mathbf{y}}\right)\|^2 \tag{3.15}$$

where the $s(\cdot)$ function operates elementwise.

Show that

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \alpha \left[(\mathbf{y} - s(\hat{\mathbf{y}})) \odot s'(\hat{\mathbf{y}}) \right] \mathbf{x}^T \tag{3.16}$$

$$\mathbf{b}(n+1) = \mathbf{b}(n) + \alpha (\mathbf{y} - s(\hat{\mathbf{y}})) \odot s'(\hat{\mathbf{y}}) \tag{3.17}$$

where $\odot$ is the Hadamard product, or elemen-twise multiplication of vectors.

3.11 Store the complete dataset in a directory and execute

```
https://raw.githubusercontent.com/gadepall/
    EE1390/master/AI−ML/codes/code.py
```

from within the directory. Note that this should be done on a powerful workstation. This will generate two files **W1.out** and **b.out**. This is the full code that is used for training.

4 Prediction

4.1 For the test input $\mathbf{x}$, compute $s(\hat{\mathbf{y}})$ using the estimated $\mathbf{W}$ and $\mathbf{b}$. If

$$\hat{y}_i = \max_j \hat{y}_j \tag{4.1}$$

define

$$y_i = 1, y_j = 0, i \neq j \tag{4.2}$$

Match this vector the entries in Table 2.2.

4.2 Execute

```
https://raw.githubusercontent.com/gadepall/
    EE1390/master/AI−ML/codes/record.py
```

to predict the output.

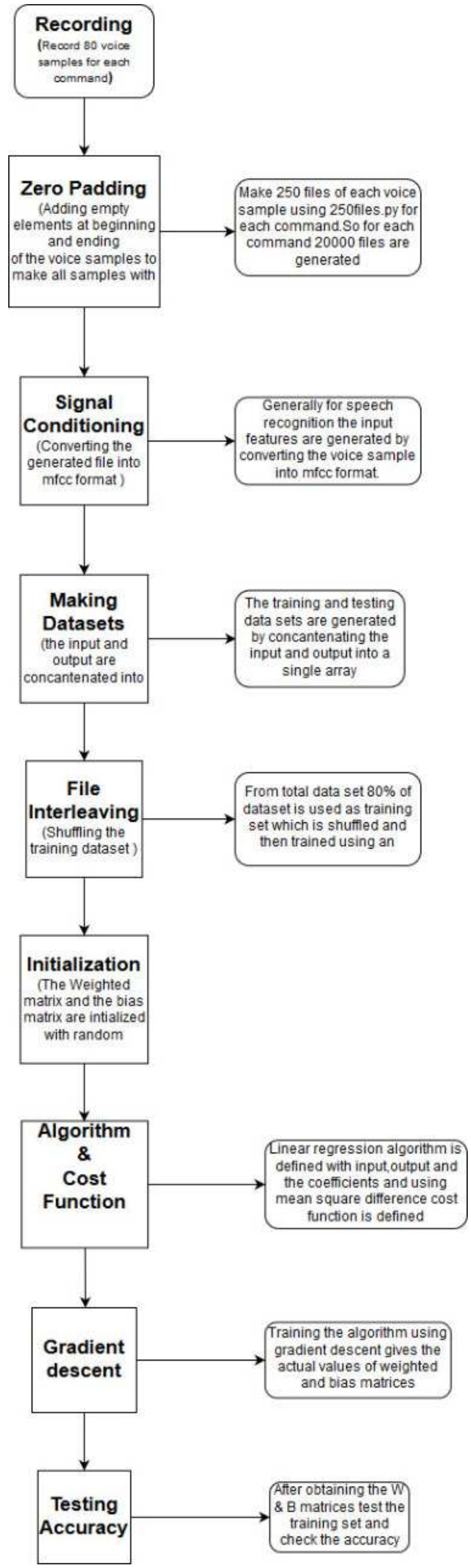

Fig. 1.1: ML System

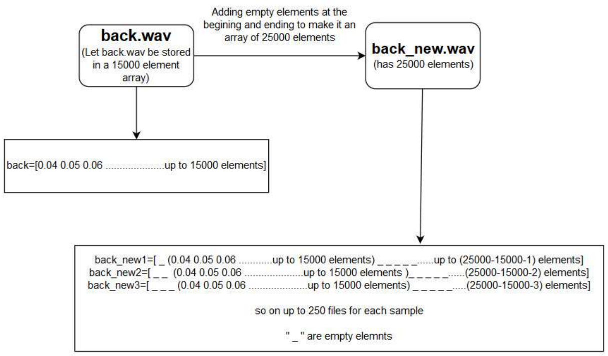

Fig. 1.4: Zero padding

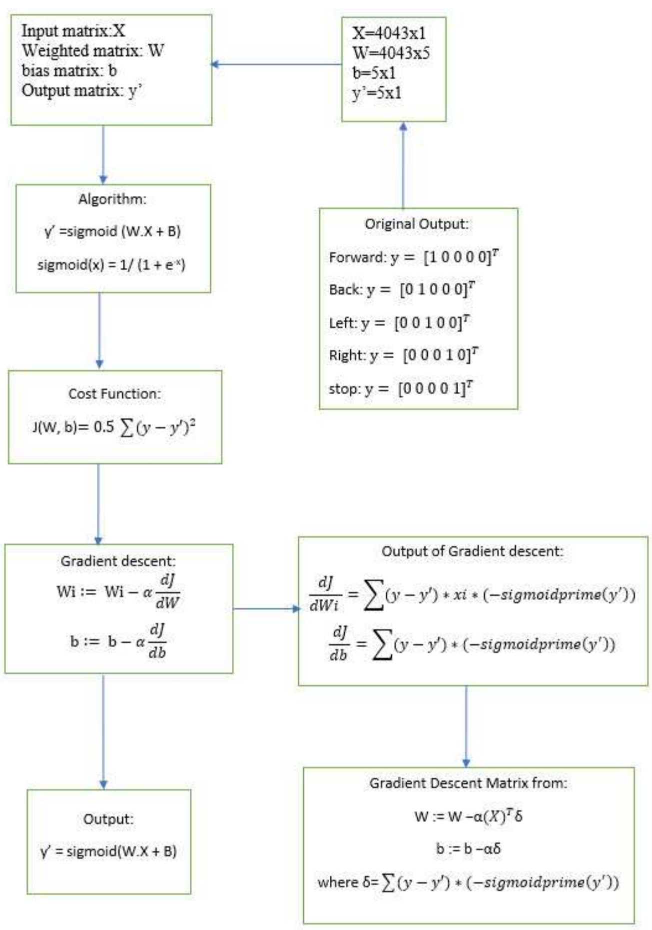

Fig. 2.1: Least squares and gradient descent

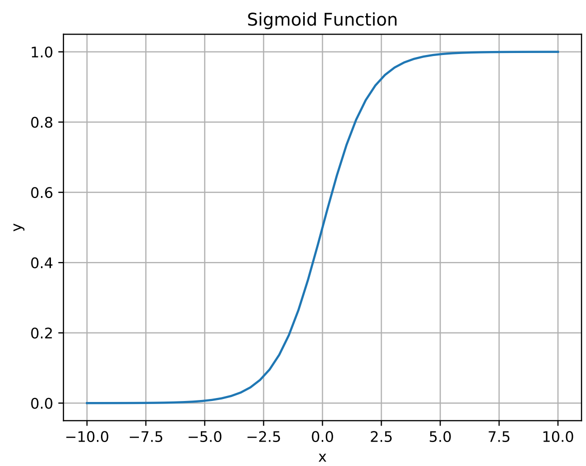

Fig. 2.3: Sigmoid function