

Contents

1	Browser Based Models with TensorFlow.js	2
1.1	Training and Inference using Tensorflow.js in JavaScript	2
1.1.1	Getting Your System Ready	2
1.1.2	API Stack	2
1.1.3	Building a First Model	2
1.1.4	Training the Model	3
1.2	Training Models with CSV Files	4
1.2.1	Reading the Data	4

1 Browser Based Models with TensorFlow.js

1.1 Training and Inference using Tensorflow.js in JavaScript

We typically think of training a neural network using GPU's or a large data center, but modern web browsers have come a long way: they contain fully fleshed out runtime environments. One of the exciting aspects of Tensorflow.js is that it allows us to train neural networks and perform inference directly from a browser. We'll see how, for example, a user can upload a picture or grab a snapshot from a webcam and then train a neural network and/or perform inference right in the browser, without ever needing to send that image up to the cloud to be processed by a server. This saves time as we cut down on communication costs, allows us to run our models offline, and preserves user privacy.

1.1.1 Getting Your System Ready

We're going to learn how to run all the examples and exercises for this course locally on a single machine. We'll use Chrome for a browser, Brackets for an HTML editor, and the Web Server for Chrome App as our web server. We will also use github.com/lmoroney/dlaicourse as a repository to store homework assignments.

1.1.2 API Stack

- At the highest level, we have the *Keras Model* layers API which we've learned how to use in the deep learning specialization.
- Beneath this sits the *Core API*, which is backed by a TensorFlow saved model. Using this Core API, we can interact with either a browser or node.js, and this is critically what allows us to use things like layers from keras.
 - A browser sits ontop of webgl, a JavaScript API for rendering graphics; we implicitly have access to a GPU here!
 - *Node.js* can rely on a TensorFlow CPU, GPU, or TPU.

1.1.3 Building a First Model

Let's start by creating the simplest possible web-page.

```
<html>
<head></head>
<body>
  <h1>First HTML Page</h1>
</body>
</html>
```

We'll next need to add a *script tag* below the head and above the body to load the TensorFlow.js file.

```
<script src="https://cdn.jsdelivr.net/npm/@tensorflow/tfjs@latest"></script>
```

We're going to build a model that learns the relationship between two numbers x, y where their ground truth relationship is $y = 2x - 1$. We will do this in a *separate script block*, that needs to be placed above the `body` tag in your HTML page.

```

<script lang="js">
  const model = tf.sequential();
  model.add(tf.layers.dense({units: 1, inputShape: [1]}));
  model.compile({loss: 'meanSquaredError',
                 optimizer: 'sgd'});
  model.summary();
</script>

```

The first line defines a *sequential* model. The second line adds a single hidden layer, itself containing a single hidden neuron. We then compile the model using mean-squared-error as our loss function, which will work well to model a linear relationship $y = 2x - 1$, and we choose stochastic gradient descent as our method of optimization. Note that in the model summary, we'll be told there are two parameters in the model, since we are learning both a weight *and* a bias term. Before closing the script tag, let's insert some data that will be used to train the neural network.

```

const xs = tf.tensor2d([-1.0, 0.0, 1.0, 2.0, 3.0, 4.0], [6, 1]);
const ys = tf.tensor2d([-3.0, -1.0, 2.0, 3.0, 5.0, 7.0], [6, 1]);

```

Notice that we're defining this as a `tensor2d`, since we don't have something like `numpy` from Python. As its name suggests, we must specify the extents of the two dimensions via the second argument.

1.1.4 Training the Model

Training should be *asynchronous*, because it takes an indeterminate amount of time to complete. Our next piece of code will call an asynchronous function called `doTraining`, which, when it completes execution will do something. Because training can take an indeterminate amount of time, we don't want to block the browser while this is taking place, so instead we specify it as an asynchronous function that calls us back when it's done.

```

doTraining(model).then(() => {
  alert(model.predict(tf.tensor2d([10], [1, 1])));
});

```

We call the function by passing it a model we just created above; when it calls back, the model is trained and at that point we can call `model.predict()`. In this example, we're predicting \hat{y} for an input of $x = 10$. Note that we still must use a `tensor2d`, which in this case is just a scalar i.e. a tensor of dimension 1×1 . To actually define an asynchronous training function, we can do so as follows:

```

async function doTraining(model){
  const history =
    await model.fit(xs, ys,
      { epochs: 500,
        callbacks:{
          onEpochEnd: async(epoch, logs) =>{
            console.log("Epoch:" + epoch + "_Loss:" + logs.loss);
          }
        }
      }
    );
}

```

This code should be placed at the top of the script block that we've been creating. Because `doTraining` is asynchronous, we use keyword `await` to wait for the result. Note that after feeding the `xs` and `ys` as input arguments, the rest of the input is a JSON list, with each list-item denoted by a name followed by a

colon, followed by a value. For callbacks, we can specify it on the fly in the list, where the callback itself is defined as a `list` and `onEpochEnd` is a function. To be clear, we're adding a function as a list-item. In our example, upon each epoch ending, we take the epoch number and `logs` as parameters so we can print out the information to console.

Because `doTraining` is asynchronous, when we call it we'll use a `then` clause to specify what should happen upon completion of execution.

1.2 Training Models with CSV Files

The Iris Dataset We've covered a simple example where we stored the values of our data in memory. A more common scenario is when data comes in from an outside source such as a database connection or an imported dataset. One of the most common ways of getting data into an ML model, and JavaScript ones are no exception, is reading data from CSV files. TensorFlow.js provides facilities for this. We'll work with an *Iris* dataset, which contains 150 samples taken from three types of Iris flower with 50 from each type. There are four features and a label for each observation.

1.2.1 Reading the Data

How can we do this using JavaScript and TensorFlow.js. Before we code, we can examine our CSV file. Note the first line contains column headers. The features encode $\{\text{sepal}, \text{petal}\} \times \{\text{length}, \text{width}\}$. We'll start by placing an asynchronous function into a JavaScript block; it must be asynchronous because we'll be waiting for some values, e.g. while training.

```
asynch function run() {  
}
```

To load the data from a CSV, we'll use the `tf.data.csv` class to handle loading and parsing the data.

```
const csvUrl = 'iris.csv';  
const trainingData = tf.data.csv(csvUrl, {  
  columnConfigs: {  
    species: {  
      isLabel: true  
    }  
  }  
});
```

There are some important details to note. The CSV resides at a URL: we don't have the server or protocol details, which means its going to try and load it from the same directory as the web-page which is hosting the application. But, it's critical to note that we're not loading from the file system directly; it's going through the HTTP stack to get the file, so you'll need to run this code on a web server.¹ In defining our training data, we make a call to `tf.data.csv` passing the URL; since tensorflow doesn't know anything about our features or labels, and so we instruct it which is our label via list syntax.

The data are returned from `tf.data.csv` as dictionaries, and for training we'll want to convert them into arrays. We'll also one-hot encode our string labels. To create a one-hot encoding, we can do so as follows:

```
const convertedData =  
  trainingData.map(( {xs, ys} => {  
    const labels = [  
      ys.species == "setosa" ? 1 : 0,  
      ys.species == "virginica" ? 1 : 0,
```

¹What's nice about the Brackets IDE is that it has a built-in web server.

```
ys.species == "versicolor" ? 1 : 0 |  
  return{ xs: Object.values(xs), ys: Object.values(labels) };  
}).batch(10);
```

The values that weren't flagged as labels are in the `xs` data-structure. If we call `Object.values(xs)`, we get back an array-of-arrays containing their values. Each row in the data-set had four features which yields a 4×1 array. These are then loaded into an array of length the size of the dataset (in this case 150 examples). Notice that we're also calling `Object.values(labels)`, which returns an array-of-arrays back as well (in this case each label is a 3×1 array and we'll have 150 of them). Ultimately, our function returns a set of features that we'll train on alongside one-hot encoded labels.