



MANGO  
SOLUTIONS

# Deep Learning in R with Keras and Google Cloud ML Engine

Doug Ashton

Twitter: @mangothecat

Email: [training@mango-solutions.com](mailto:training@mango-solutions.com)



# Agenda

- Introduction to Deep Learning
- Google Cloud Platform (GCP)
- CloudML
  - Built in data
  - Supplied data
  - Cloud storage



# About Me

- Physics



Started at Balliol

- 10 years research
- Soft Matter / Monte Carlo / Graphs

- Mango:



Data Scientist at Mango 5 years

- Consulting / Training / R community



@dougashton

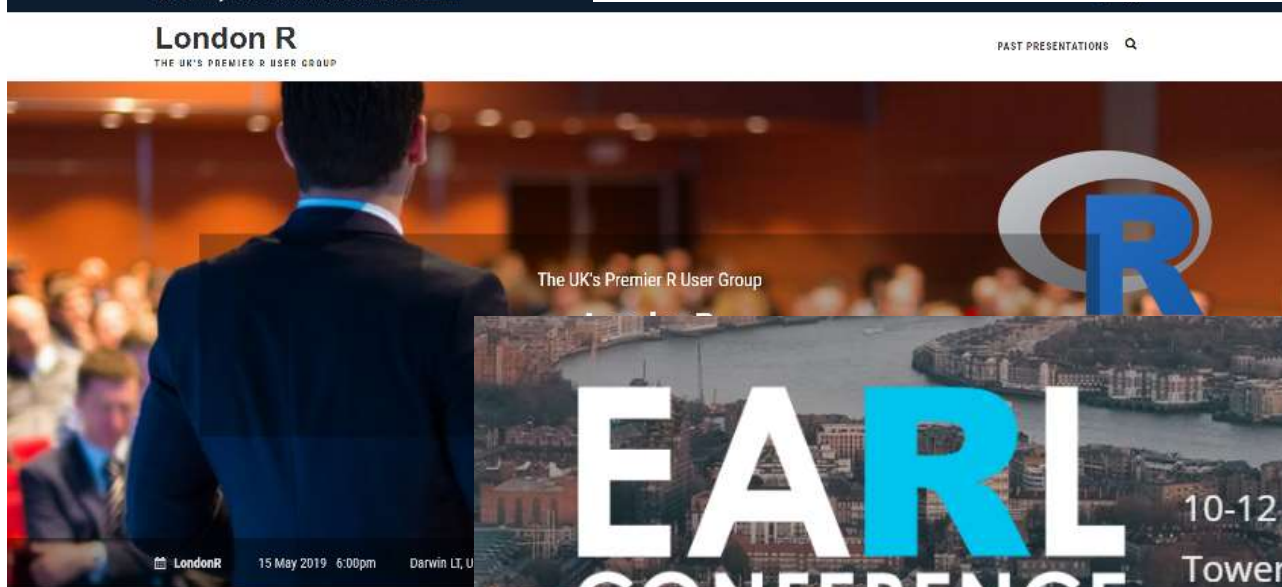


dougmet





Code: <https://github.com/dougmet/cloudml>



[londonr.org](http://londonr.org)

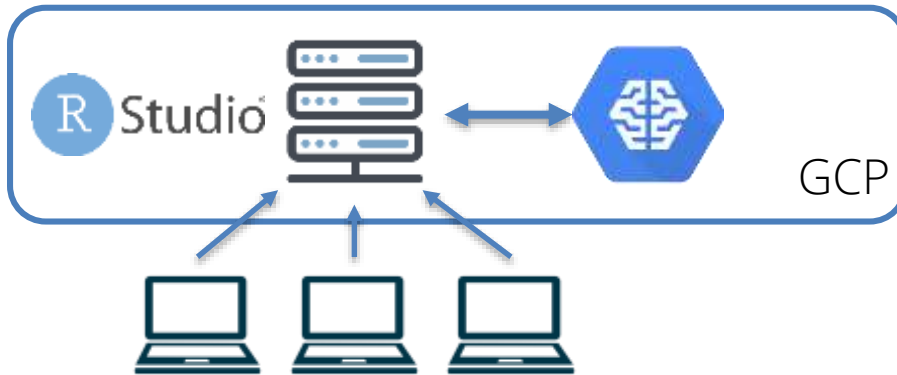


[earlconf.com](http://earlconf.com)

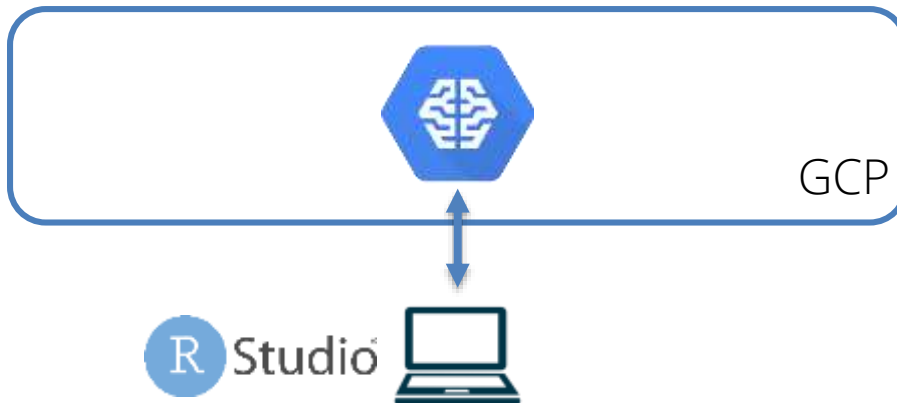


Code: <https://github.com/dougmet/cloudml>

# CloudML Use Cases



Cloud project. Central RSS.  
Run big jobs in ML-engine



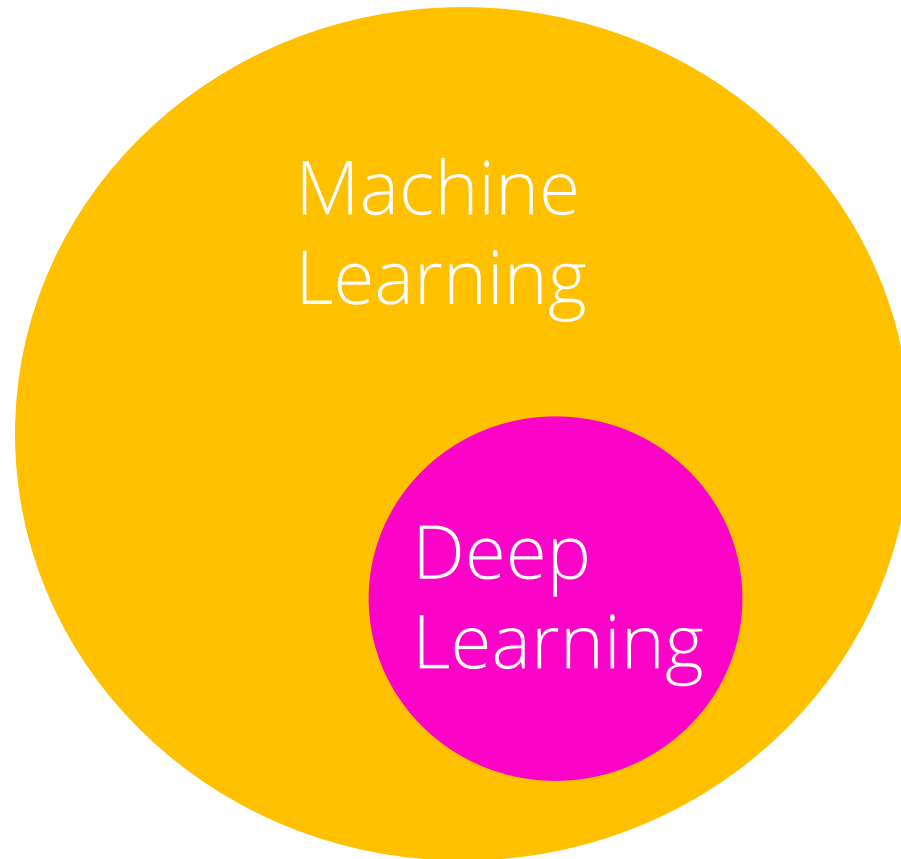
Local project.  
Run big jobs in ML-engine



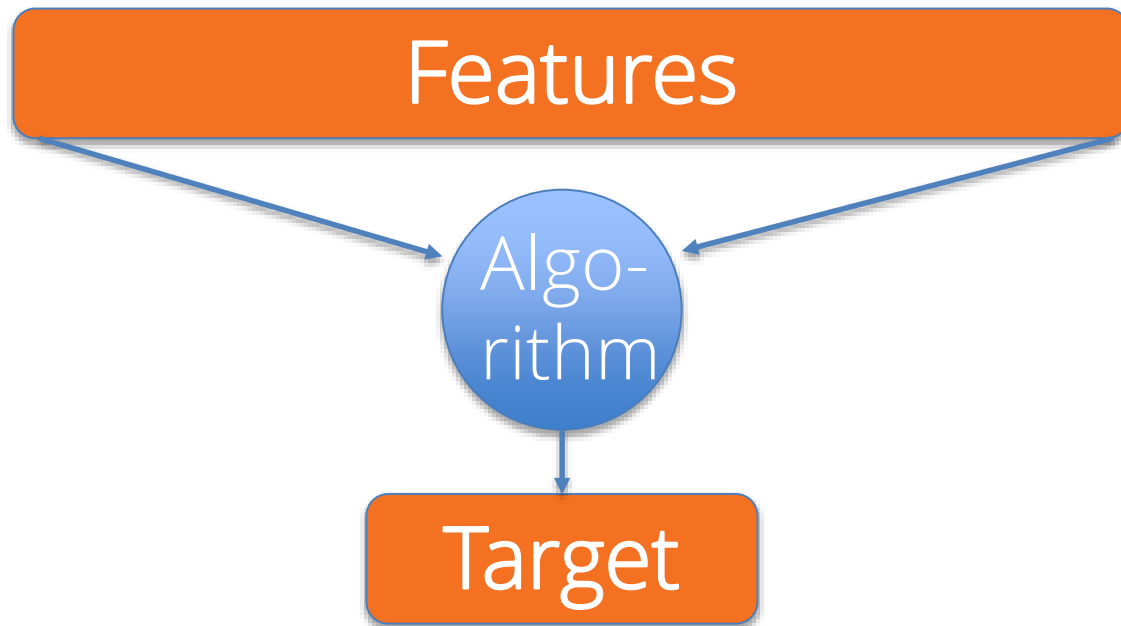
# Introduction to Deep Learning



# What is Deep Learning?



# What is Deep Learning?

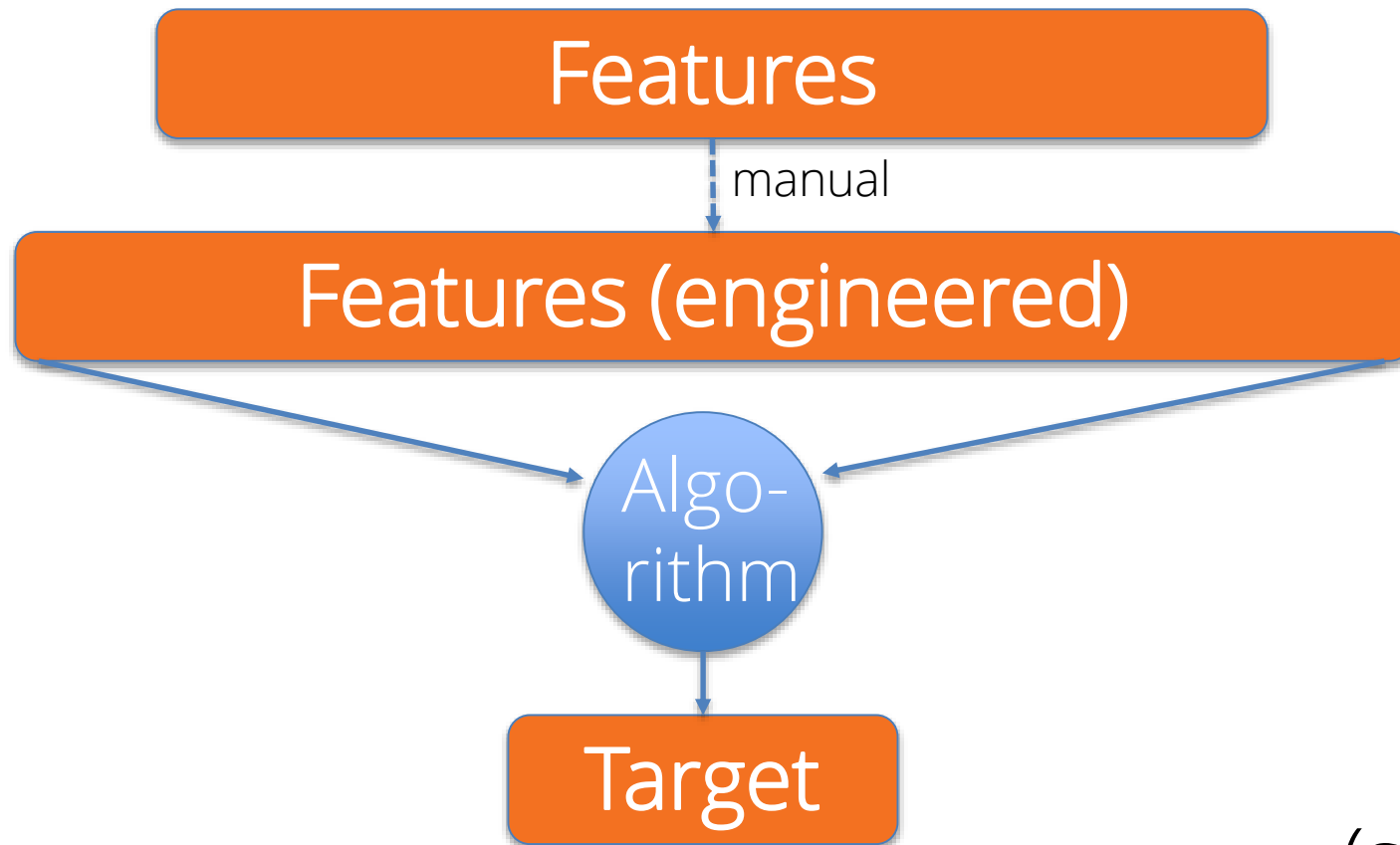


(shallow)





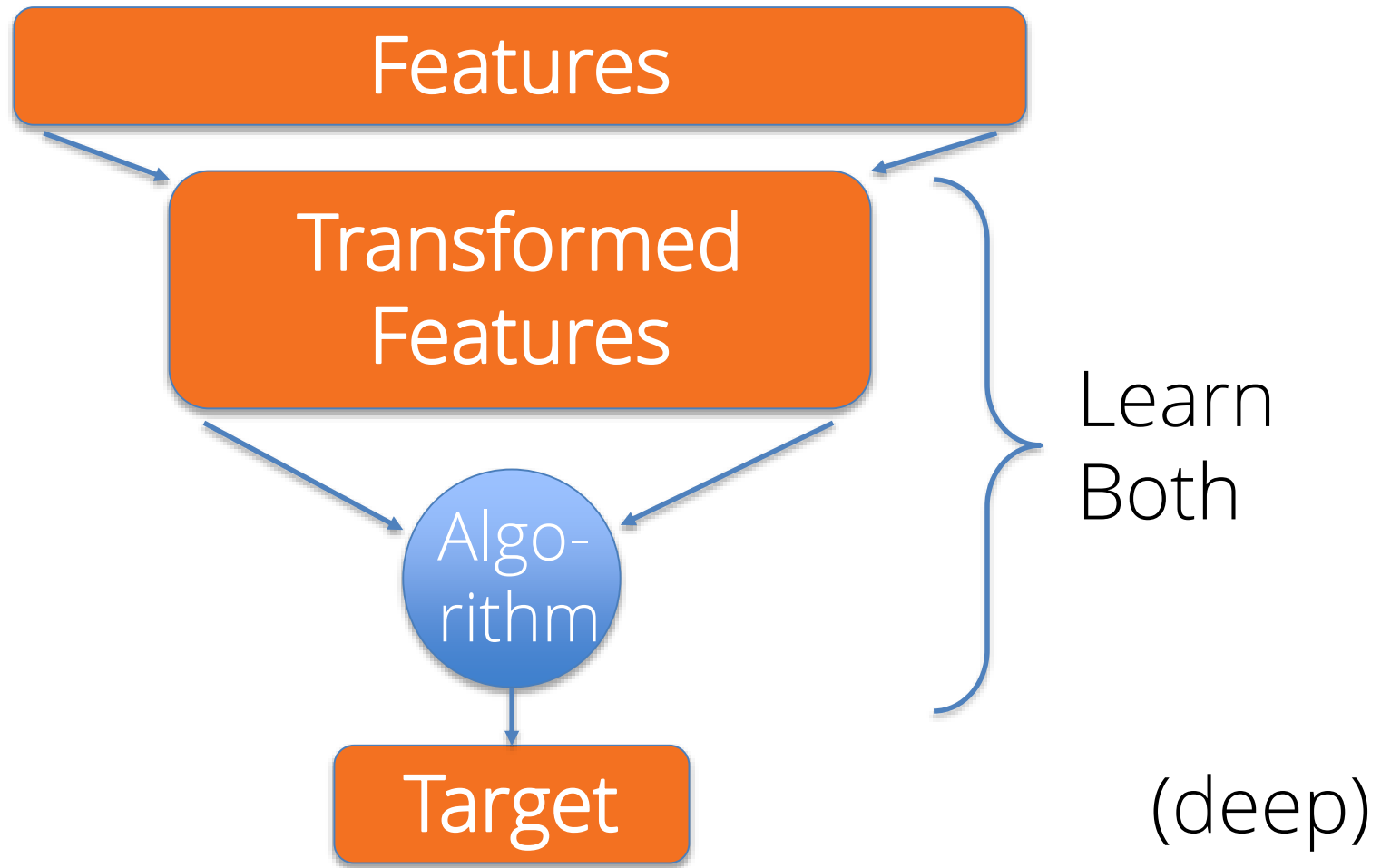
# What is Deep Learning?



(shallow)



# What is Deep Learning?



# What Does it Solve?

- Unstructured
  - Features are learned rather than designed
- Big
  - Generally need lots of data
- Familiar
  - Can reuse models on new problems



# Spatial

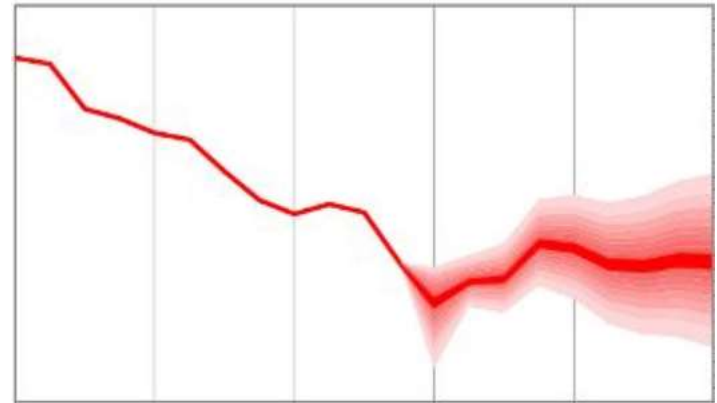
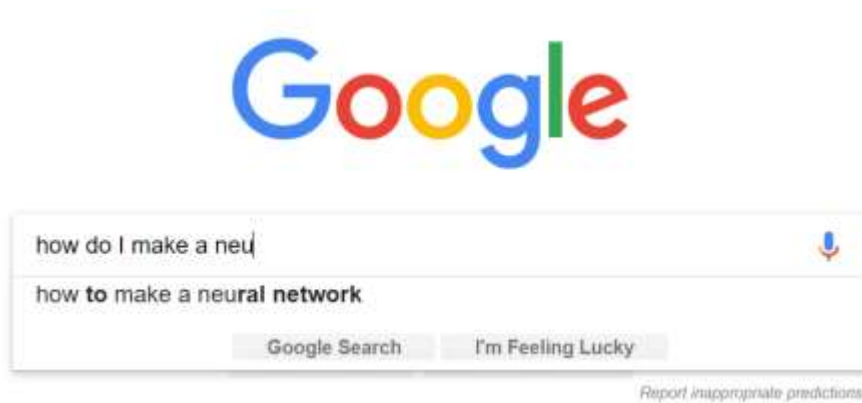
- Computer vision
- Audio
- Time series: pattern recognition

IMGENET



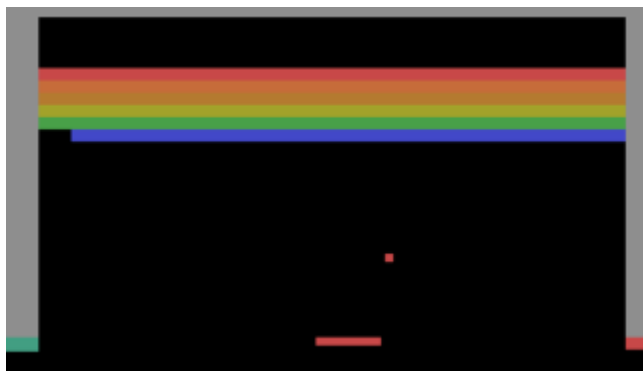
# Sequential

- Language
- Time series: Forecasting



# Reinforcement/Adversarial

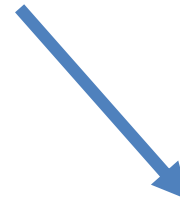
- AlphaGo
- Generative Networks



# Neural Networks



nodes



edges



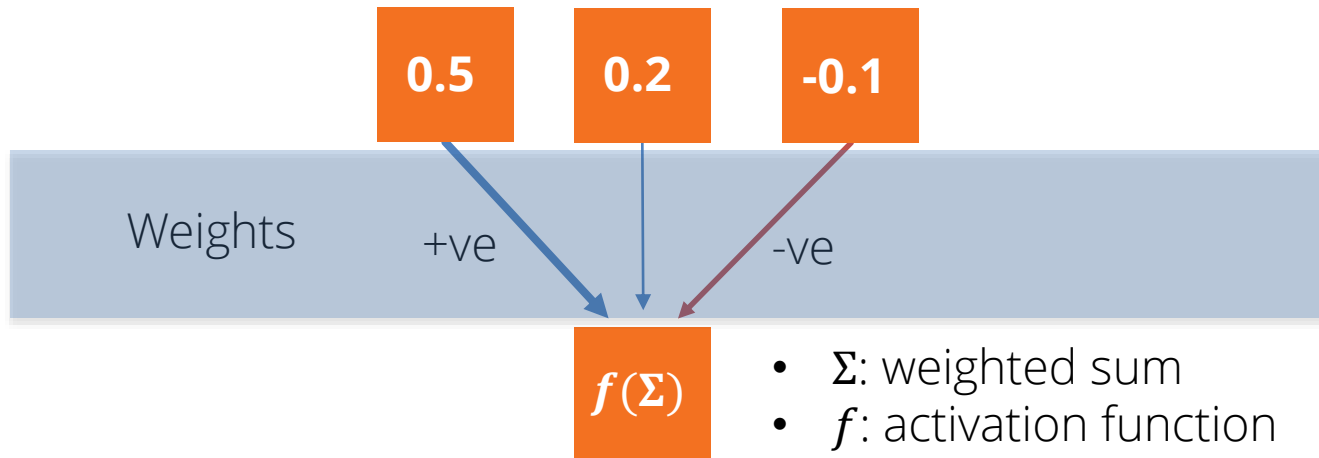
# A Neuron

0.2





# Neurons

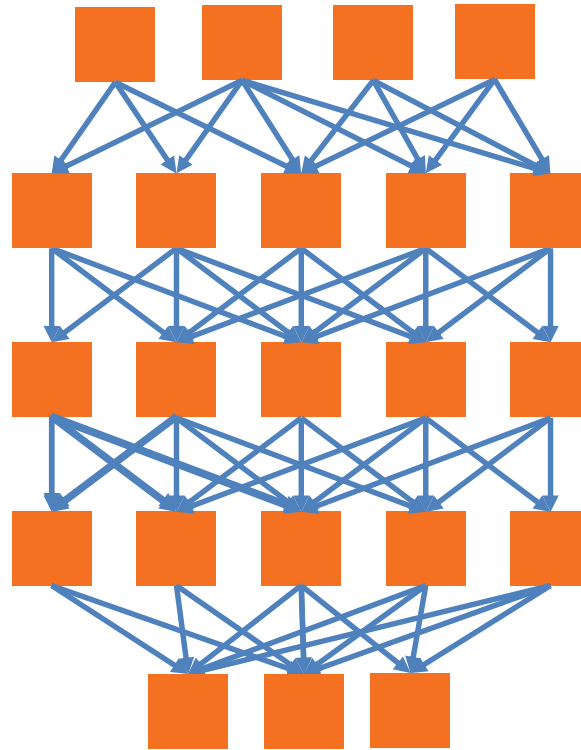


# Neural Network

Input layer

Hidden layers

Output layer



More  
abstract



# TensorFlow

- Turns equations into dataflow graphs
  - <https://www.tensorflow.org>
- Efficient numerical solver
- Built for CPU, GPU, and TPU
- Not only for neural networks



TensorFlow



# TensorFlow and R

- RStudio built an R interface
  - <https://tensorflow.rstudio.com>
- Python <-> R handled by reticulate
  - <https://rstudio.github.io/reticulate>



# Keras

- High level interface for building neural networks
  - <https://keras.io>
  - François Chollet
- Works with multiple backends
  - TensorFlow, CNTK, Theano, MXNet, CoreML,...

News! Keras is to be integrated tightly in TensorFlow 2.0

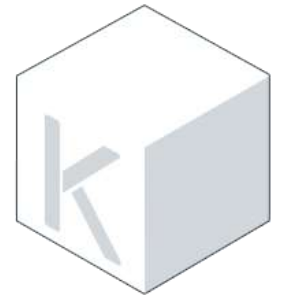


Keras



# Keras and R

- Rstudio built an interface to Keras
  - <https://keras.rstudio.com>
- Also to TensorFlow directly

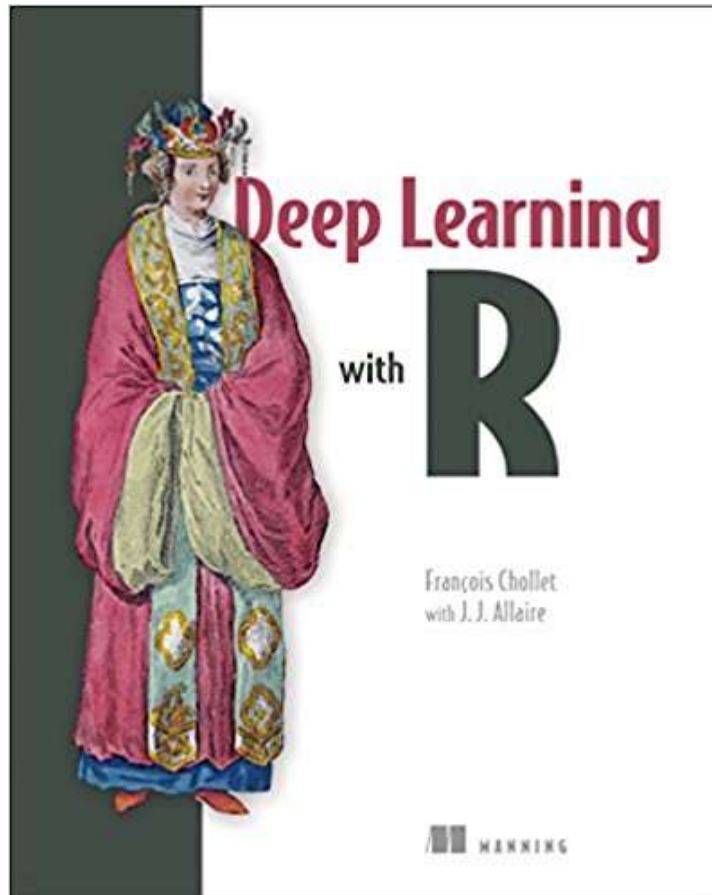


Keras API



# Keras

# Keras and R Book



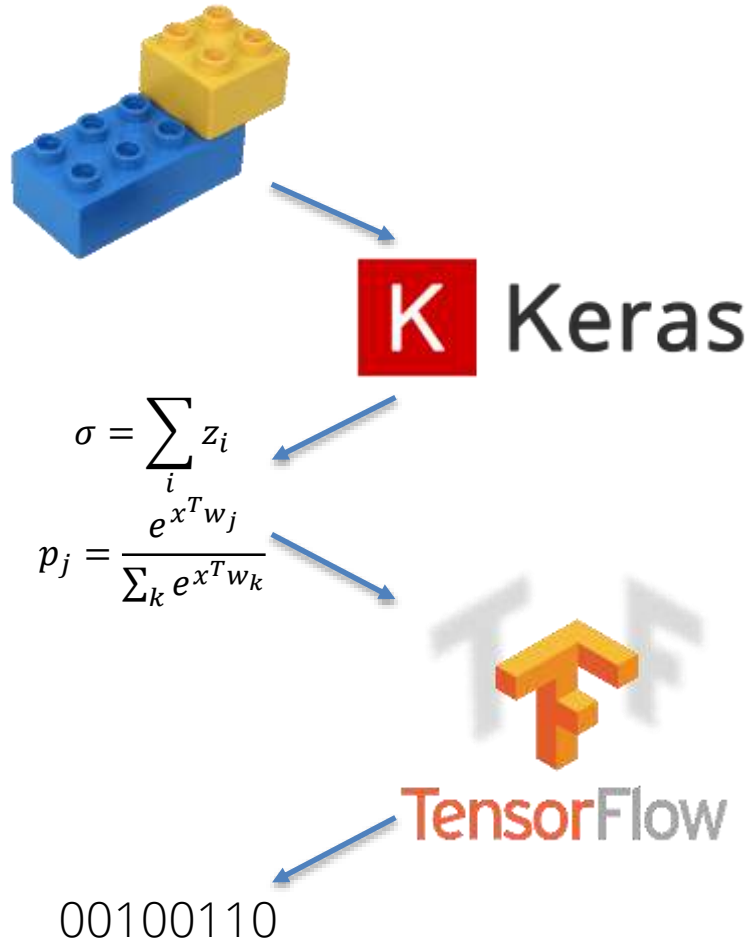
Deep Learning with R  
- François Chollet  
- J. J. Allaire

Manning

<https://www.manning.com/books/deep-learning-with-r>



# How it fits together





# Alternatives for R Users

- MXNet
  - Native R support
  - <https://mxnet.incubator.apache.org/api/r/>
  - Amazon investment  
<https://aws.amazon.com/mxnet/>





# Google Cloud Platform



# Disclosure

- Mango are Google partners
- Not exclusive but GCP is our default cloud
- AWS has equivalents for most things here



Google Cloud Platform

ML Engine ~ SageMaker  
Cloud Storage ~ S3  
BigQuery ~ RedShift



# Trends in Cloud Computing

- Separation of storage and compute
  - (no more Hadoop?)
- Service based
  - (no more servers?)



# Separate Storage and Compute

## Storage



Cloud  
Storage



BigQuery

High speed  
network

## Compute



ML Engine



Data Labs



Compute  
Engine



# Serverless

## Storage



Cloud  
Storage



BigQuery

High speed  
network

## Compute



ML Engine



Data Labs



~~Compute  
Engine~~



# Rstudio and ML Engine

- [https://tensorflow.rstudio.com/tools/cloudml/articles/getting\\_started.html](https://tensorflow.rstudio.com/tools/cloudml/articles/getting_started.html)

TensorFlow for R



## R Interface to Google CloudML

### Overview

The `cloudml` package provides an R interface to [Google Cloud Machine Learning Engine](#), a managed service that enables:

- Scalable training of models built with the [keras](#), [tfestimators](#), and [tensorflow](#) R packages.
- On-demand access to training on GPUs, including the new [Tesla P100 GPUs](#) from NVIDIA®.
- Hyperparameter tuning to optimize key attributes of model architectures in



# Should I just use Python?

+ R

- Continue R-based workflow (**any** R-based job)
- cloudml really easy →

Job ID	Type	Created
pytest000057	Training	Apr 7
pytest000058	Training	Apr 7
pytest000059	Training	Apr 7
pytest000060	Training	Apr 7
pytest000061	Training	Apr 7
pytest000062	Training	Apr 7
pytest000063	Training	Apr 7
cloudml_2019_04_05_15302934	Training	Apr 5
cloudml_2019_04_05_15302937	Training	Apr 5
cloudml_2019_04_05_09430072	Training	Apr 5
cloudml_2019_04_05_09395877	Training	Apr 5
cloudml_2019_04_05_09304381	Training	Apr 5
cloudml_2019_04_05_09175178	Training	Apr 5
cloudml_2019_04_04_22463020	Training	Apr 4



+ Python

- Native Python support
- Packages pre-installed
- More flexible in deployment

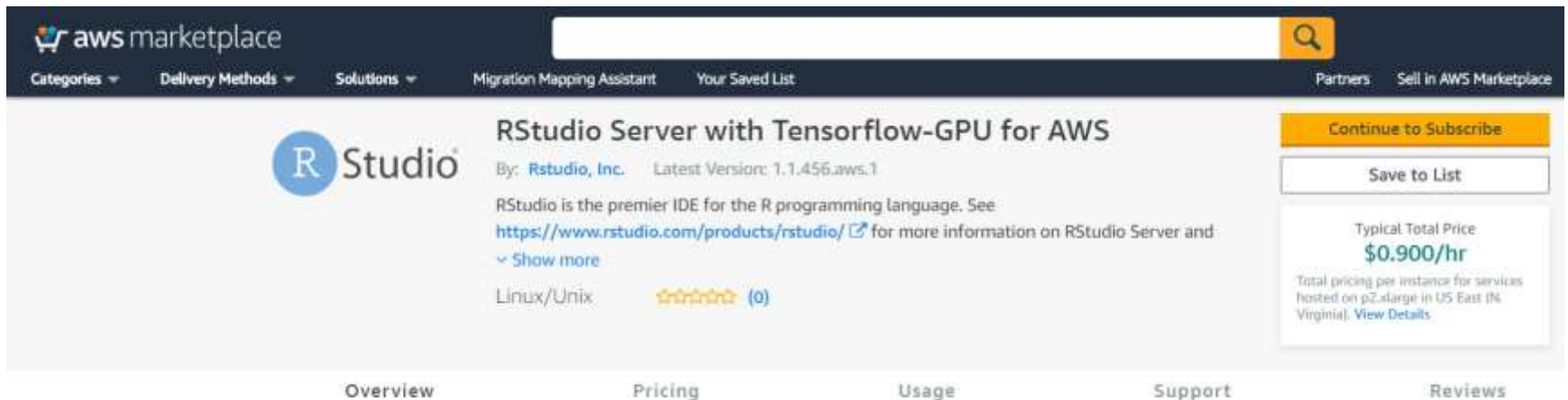
gs://keras-235720/py-cloudml  
12.74 KiB

gs://keras-235720/r-cloudml  
1.03 GiB






# AWS RStudio Server Alternative



The screenshot shows the AWS Marketplace interface for the product "RStudio Server with Tensorflow-GPU for AWS". The header includes the AWS Marketplace logo, a search bar, and navigation links for Categories, Delivery Methods, Solutions, Migration Mapping Assistant, Your Saved List, Partners, and Sell in AWS Marketplace. The product listing features the RStudio logo, the product name, the provider "Rstudio, Inc.", and the latest version "1.1.456.aws.1". A description states that RStudio is the premier IDE for the R programming language, with a link to the RStudio website for more information. Below the description, it indicates the operating system "Linux/Unix" and shows a star rating of 0 reviews. On the right side, there are buttons for "Continue to Subscribe" and "Save to List", along with a pricing box showing a "Typical Total Price" of "\$0.900/hr" and a note about total pricing per instance for services hosted on p2.xlarge in US East (N. Virginia), with a link to "View Details". At the bottom, there are tabs for "Overview", "Pricing", "Usage", "Support", and "Reviews".


aws marketplace

Categories ▾ Delivery Methods ▾ Solutions ▾ Migration Mapping Assistant Your Saved List Partners Sell in AWS Marketplace

 **RStudio**

By: [Rstudio, Inc.](#) Latest Version: 1.1.456.aws.1

RStudio is the premier IDE for the R programming language. See <https://www.rstudio.com/products/rstudio/> for more information on RStudio Server and [Show more](#)

Linux/Unix  (0)

[Continue to Subscribe](#)

[Save to List](#)

Typical Total Price  
**\$0.900/hr**  
Total pricing per instance for services hosted on p2.xlarge in US East (N. Virginia). [View Details](#)

Overview Pricing Usage Support Reviews



# Setting it Up



# Make a Project

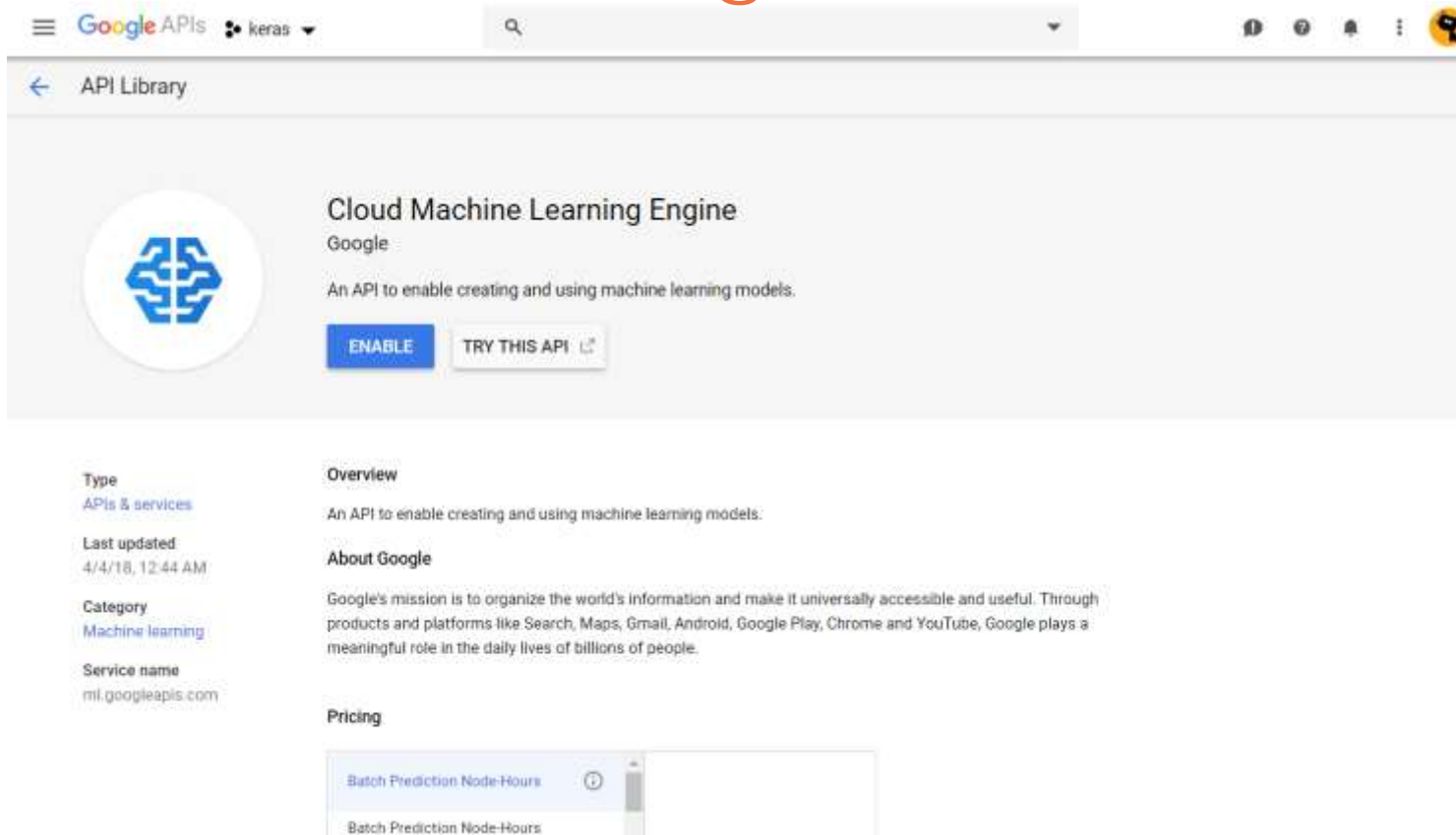
The screenshot shows the Google Cloud Platform (GCP) dashboard for a project named 'keras'. The interface is organized into several sections:

- Project info:** Displays the project name 'keras', Project ID 'keras-235720', and Project number '372751896196'. A link 'Go to project settings' is provided.
- Resources:** A section indicating 'This project has no resources'.
- Trace:** A section indicating 'No trace data from the past 7 days' with a link 'Get started with Stackdriver Trace'.
- RPI APIs:** A section showing a graph of 'Requests (requests/sec)' over time. The graph is empty, with a message: 'No data is available for the selected time frame'. A link 'Go to APIs overview' is provided.
- Google Cloud Platform status:** A section showing 'All services normal' and a link 'Go to Cloud status dashboard'.
- Billing:** A section showing 'Estimated charges' of 'GBP £0.00' for the billing period 'Mar 1 - 26, 2019'. A link 'View detailed charges' is provided.
- Error Reporting:** A section showing 'No sign of any errors. Have you set up Error Reporting?' and a link 'Learn how to set up Error Reporting'.
- News:** A section for news updates.

The top navigation bar includes the Google Cloud Platform logo, the project name 'keras', a search bar, and various utility icons. The left sidebar shows 'DASHBOARD' and 'ACTIVITY' tabs.




# Enable ML Engine API



The screenshot shows the Google APIs Library interface. At the top, there's a navigation bar with the Google APIs logo, a search bar, and user avatars. Below the navigation bar, the page title is "API Library". The main content area features the "Cloud Machine Learning Engine" API by Google. It includes a circular icon with a blue brain-like pattern. The text describes it as "An API to enable creating and using machine learning models." There are two buttons: "ENABLE" in blue and "TRY THIS API" in white with a link icon. Below this, there are sections for "Type" (APIs & services), "Last updated" (4/4/18, 12:44 AM), "Category" (Machine learning), and "Service name" (ml.googleapis.com). The "Overview" section repeats the description. The "About Google" section mentions Google's mission. The "Pricing" section shows a table with one row: "Batch Prediction Node-Hours".

Google APIs keras

API Library

 **Cloud Machine Learning Engine**  
Google

An API to enable creating and using machine learning models.

**ENABLE** TRY THIS API

**Type**  
APIs & services

**Last updated**  
4/4/18, 12:44 AM

**Category**  
Machine learning

**Service name**  
ml.googleapis.com

**Overview**  
An API to enable creating and using machine learning models.

**About Google**  
Google's mission is to organize the world's information and make it universally accessible and useful. Through products and platforms like Search, Maps, Gmail, Android, Google Play, Chrome and YouTube, Google plays a meaningful role in the daily lives of billions of people.

**Pricing**

Batch Prediction Node-Hours
Batch Prediction Node-Hours



# Install SDK

```
install.packages("cloudml")
```

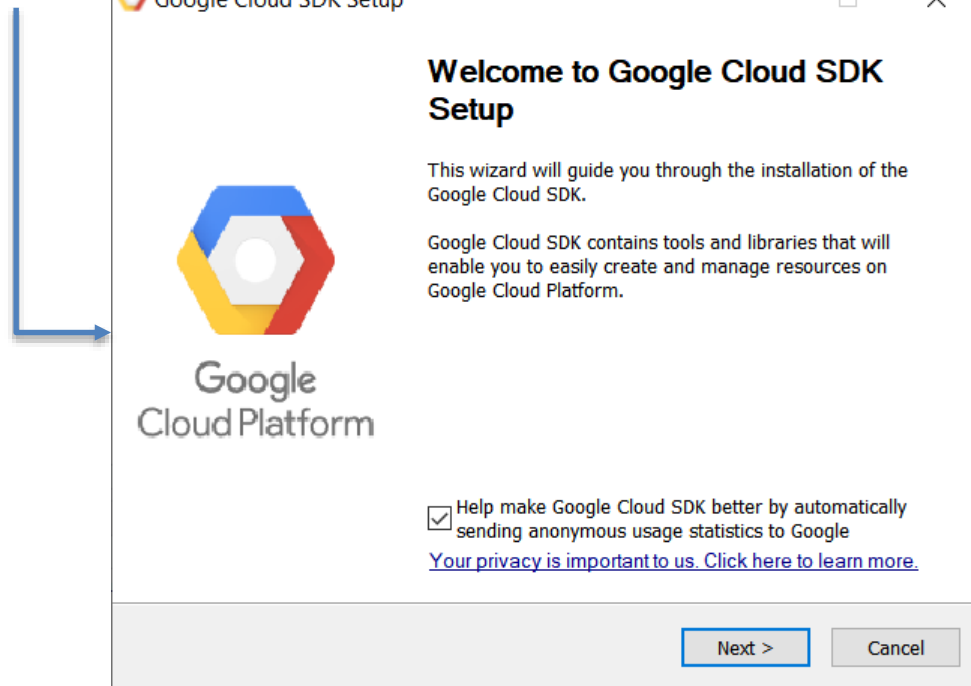
```
library(cloudml)
```

```
gcloud_install()
```

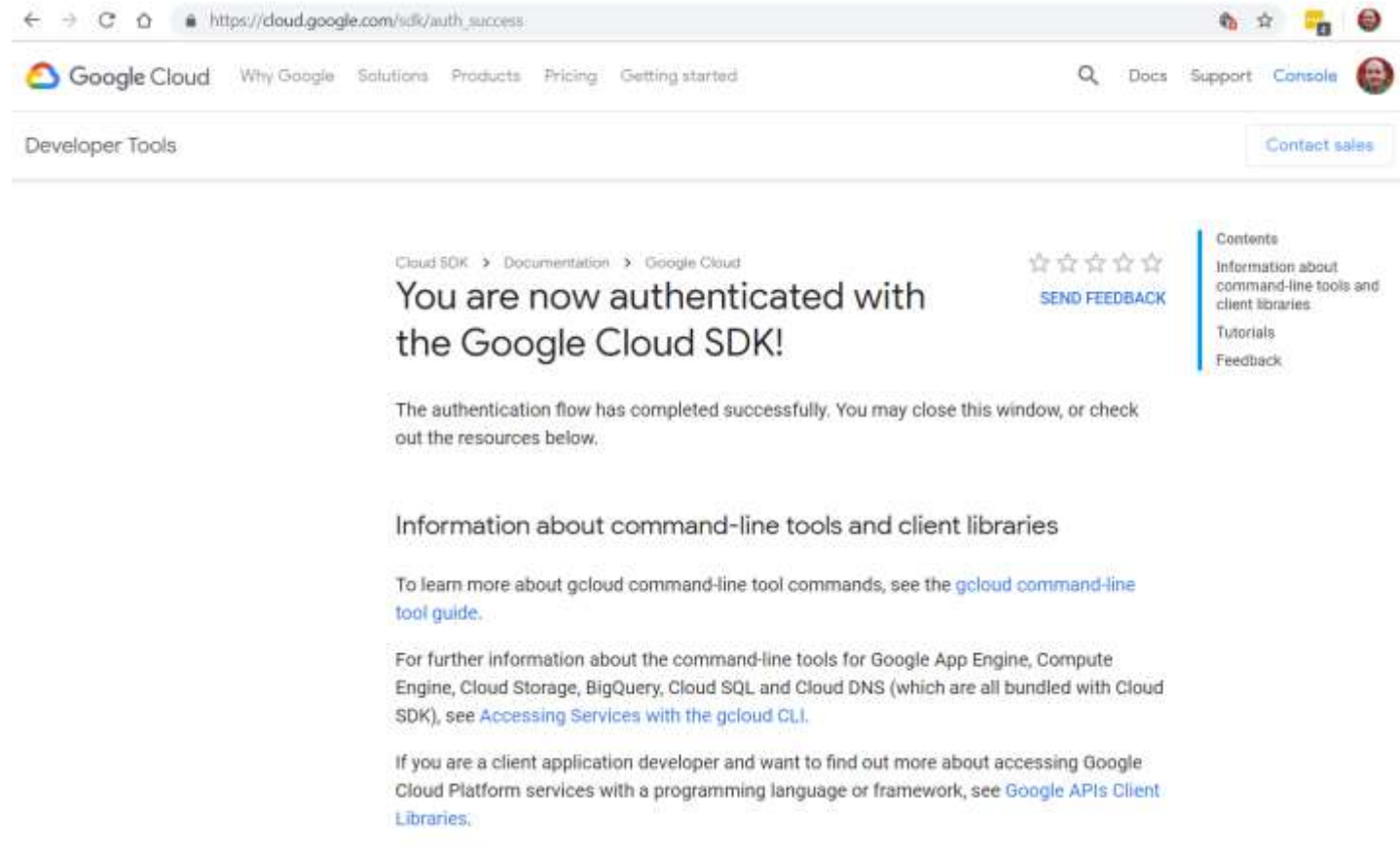
```
# I'm taking the default location in my user dir.
```

```
# It didn't ask me about projects so running init again
```

```
gcloud_init()
```



# Login



The screenshot shows a web browser window with the URL [https://cloud.google.com/sdk/auth\\_success](https://cloud.google.com/sdk/auth_success). The page header includes the Google Cloud logo and navigation links: Why Google, Solutions, Products, Pricing, and Getting started. On the right, there are links for Docs, Support, and Console, along with a user profile icon. Below the header, the text "Developer Tools" is on the left and a "Contact sales" button is on the right. The main content area features a breadcrumb trail: Cloud SDK > Documentation > Google Cloud. The primary heading is "You are now authenticated with the Google Cloud SDK!". To the right of this heading are five stars and a "SEND FEEDBACK" link. Below the heading, a message states: "The authentication flow has completed successfully. You may close this window, or check out the resources below." A section titled "Information about command-line tools and client libraries" follows. It contains two paragraphs: the first points to the "gcloud command-line tool guide" for more about gcloud commands; the second provides further information about command-line tools for Google App Engine, Compute Engine, Cloud Storage, BigQuery, Cloud SQL, and Cloud DNS, all bundled with the Cloud SDK, and refers to "Accessing Services with the gcloud CLI". A final paragraph addresses client application developers, directing them to "Google APIs Client Libraries" for more on accessing Google Cloud Platform services with a programming language or framework. A right-hand sidebar contains a "Contents" section with links to "Information about command-line tools and client libraries", "Tutorials", and "Feedback".

Cloud SDK > Documentation > Google Cloud

## You are now authenticated with the Google Cloud SDK!

★★★★★  
[SEND FEEDBACK](#)

The authentication flow has completed successfully. You may close this window, or check out the resources below.

### Information about command-line tools and client libraries

To learn more about gcloud command-line tool commands, see the [gcloud command-line tool guide](#).

For further information about the command-line tools for Google App Engine, Compute Engine, Cloud Storage, BigQuery, Cloud SQL and Cloud DNS (which are all bundled with Cloud SDK), see [Accessing Services with the gcloud CLI](#).

If you are a client application developer and want to find out more about accessing Google Cloud Platform services with a programming language or framework, see [Google APIs Client Libraries](#).

**Contents**  
[Information about command-line tools and client libraries](#)  
[Tutorials](#)  
[Feedback](#)



# Maybe run init (Rstudio)

Had to run gcloud\_init again to pickup projects

```
C:\Users\dashton\Documents\R\cloudml>"C:\Users\dashton\AppData\Local\Google\
Welcome! This command will take you through the configuration of gcloud.
```

```
Settings from your current configuration [default] are:
core:
```

```
  account: dashton@mango-solutions.com
  disable_usage_reporting: 'False'
```

```
Pick configuration to use:
```

- [1] Re-initialize this configuration [default] with new settings
- [2] Create a new configuration

```
Please enter your numeric choice: 1
```





# Choose project

this configuration:

```
[1] dashton@mango-solutions.com  
[2] Log in with a new account  
Please enter your numeric choice: 1
```

You are logged in as: [dashton@mango-solutions.com].

Pick cloud project to use:

```
[1]   
[2]   
[3] keras-235720  
[4] mango-233109  
[5]   
[6] Create a new project  
Please enter numeric choice or text value (must exactly match list  
item): 3
```

Your current project has been set to: [keras-235720].





# First Keras Model





Code: <https://github.com/dougmet/cloudml>

# Iris Neural Network

`iris[1,1:4]`

Sepal.Length	Sepal.Width	Petal.Width	Petal.Length
5.1	3.5	1.4	0.2

Features  $\mathbf{x}$

`iris[1,5]`

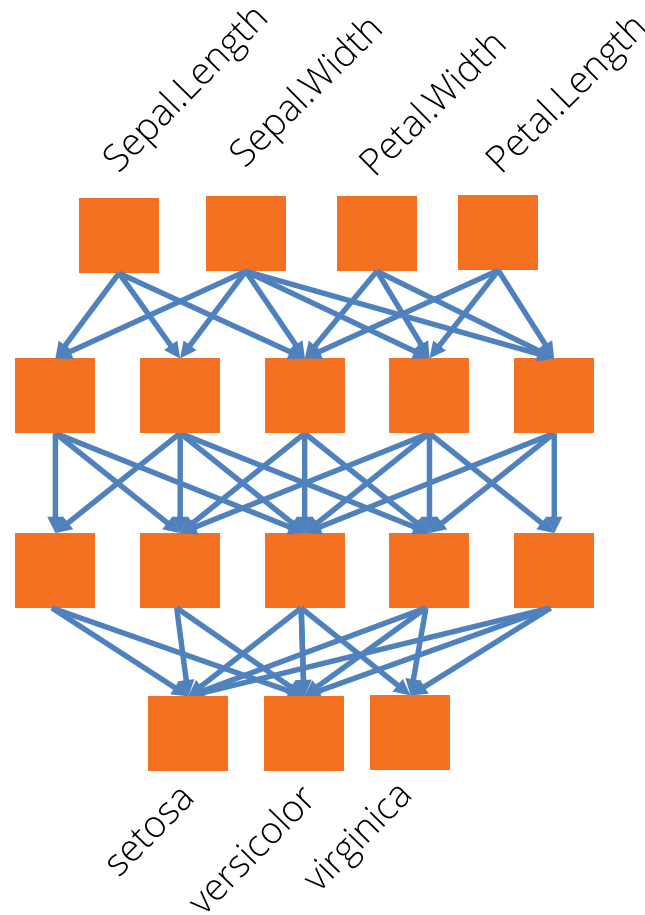
1	0	0
setosa	versicolor	virginica

Target  $\mathbf{y}$



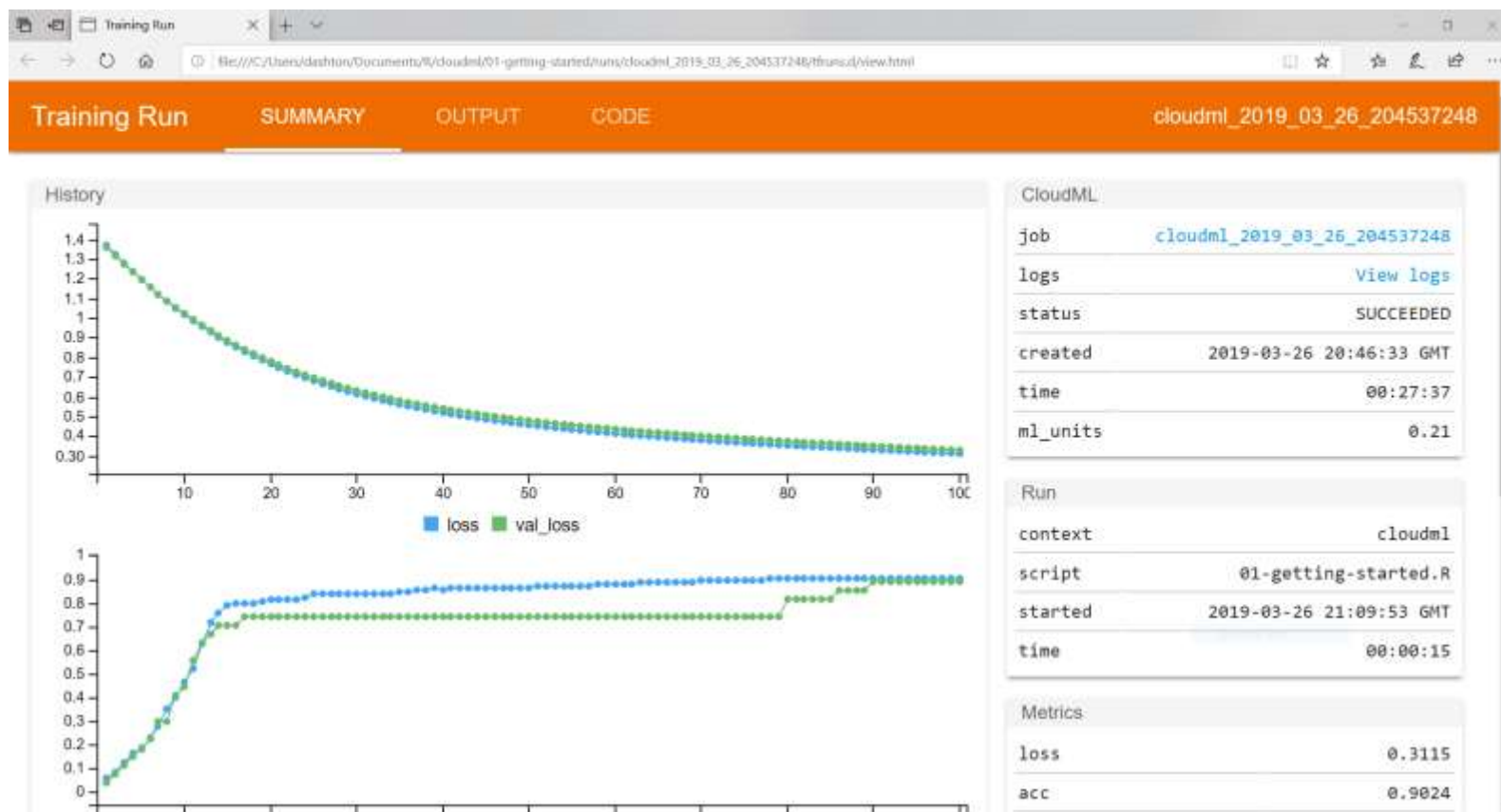
Code: <https://github.com/dougmet/cloudml>

# Iris Neural Network



Code: <https://github.com/dougmet/cloudml>

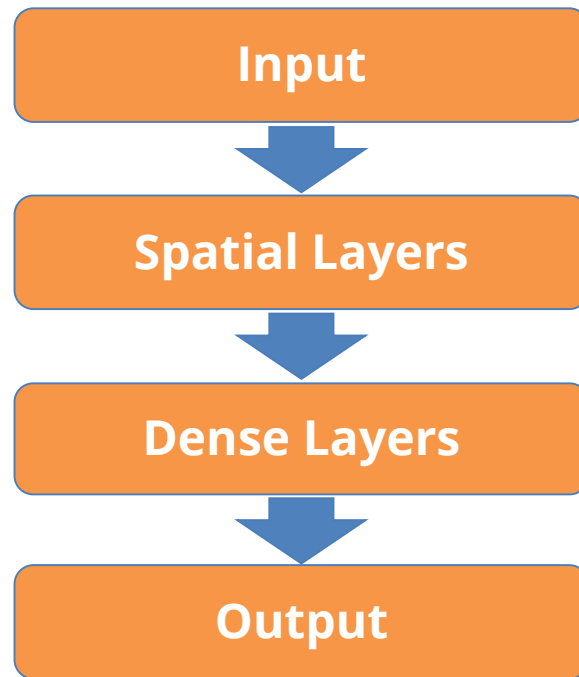
# Demo!



# Networks for Spatial Data



# Convolutional Neural Networks



# Walking Data

- Accelerometer data from the UCI
- Filtered to walking activity
- 15 Different people
- Can we recognise someone by their gait?
- Chopped into 5 second chunks

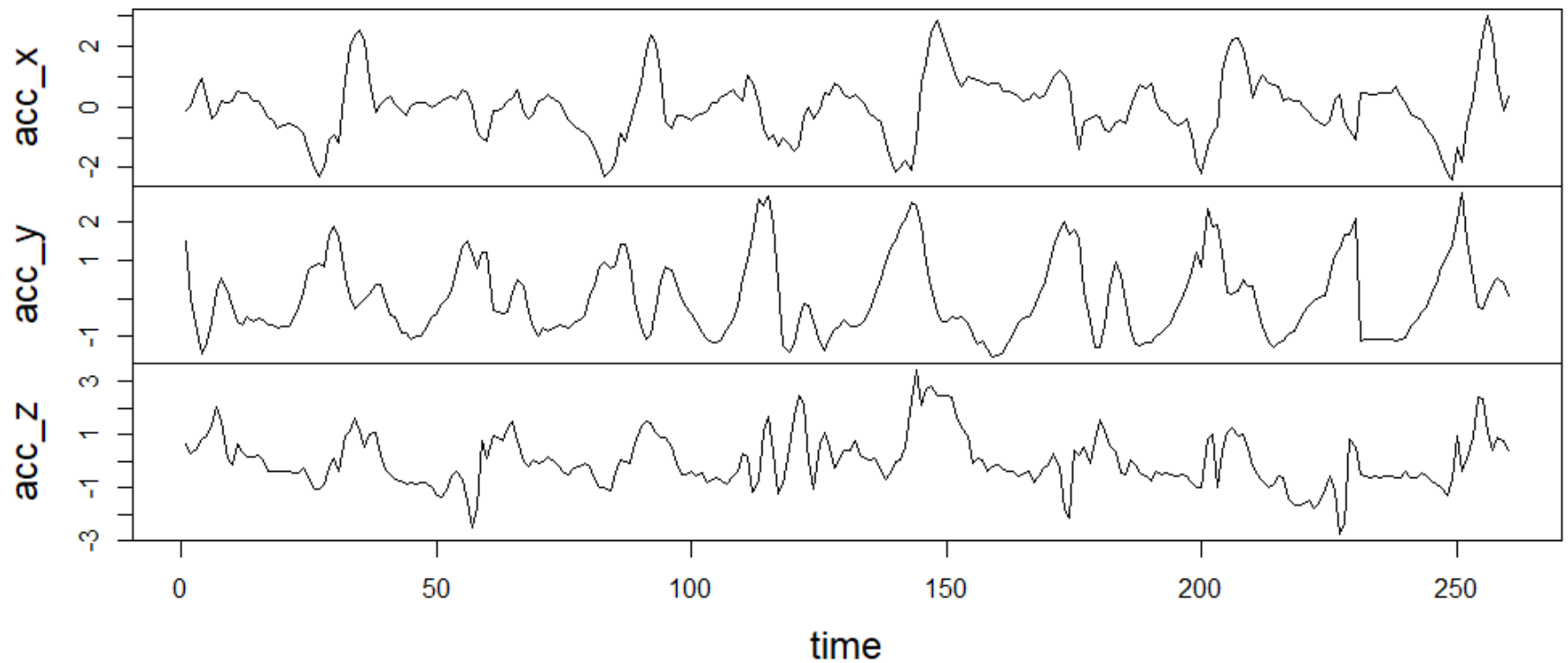
<https://archive.ics.uci.edu/ml/datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer>



# Walking Data

Walking[50,,]

**Person 1**

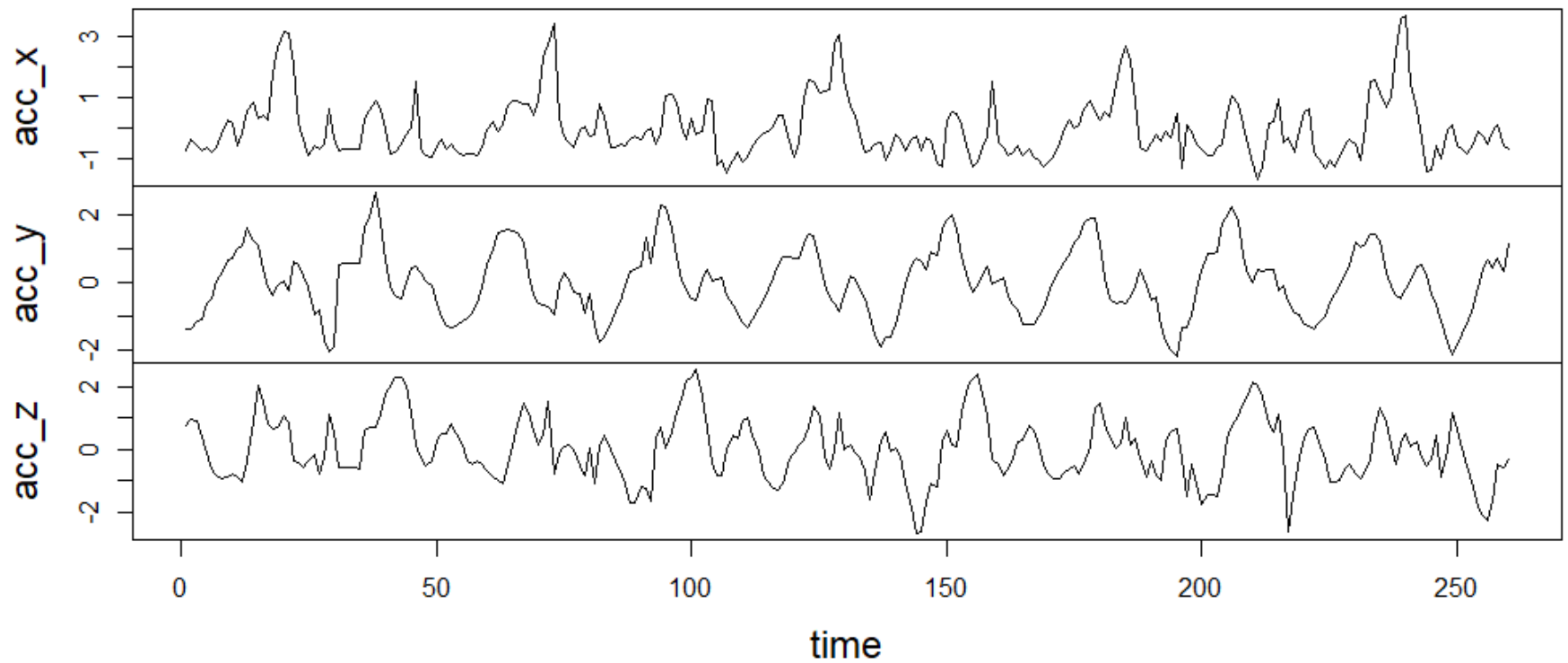




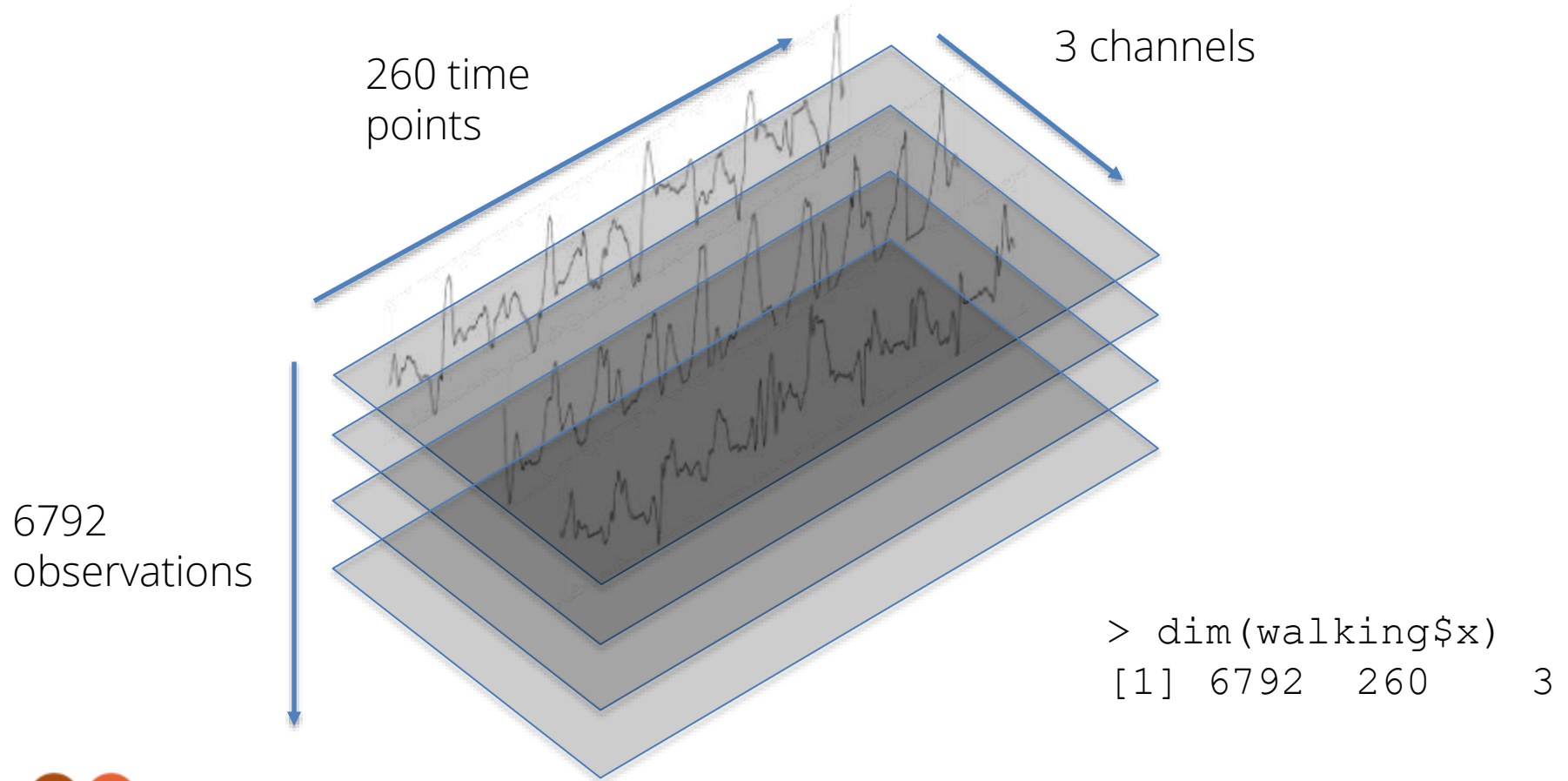
# Walking Data

Walking[4100,,]

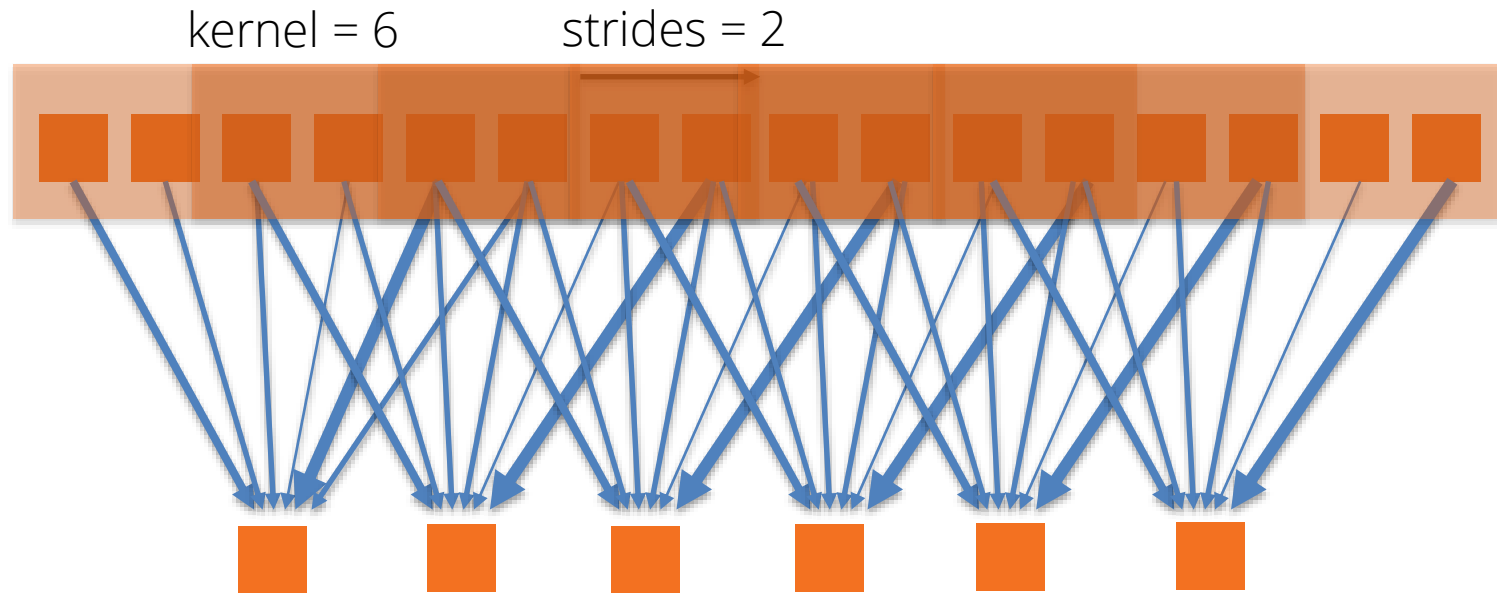
**Person 10**



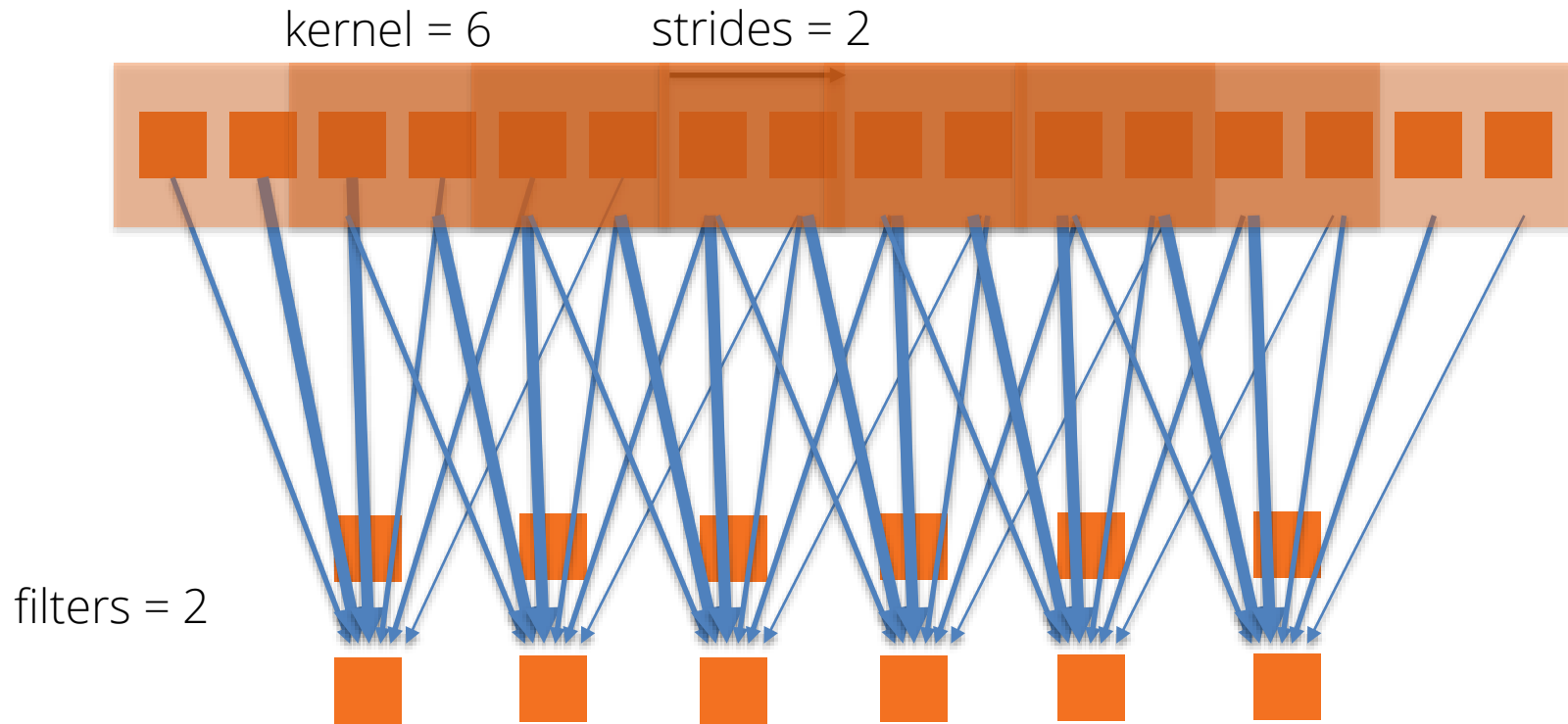
# Walking Data



# Convolution Layer

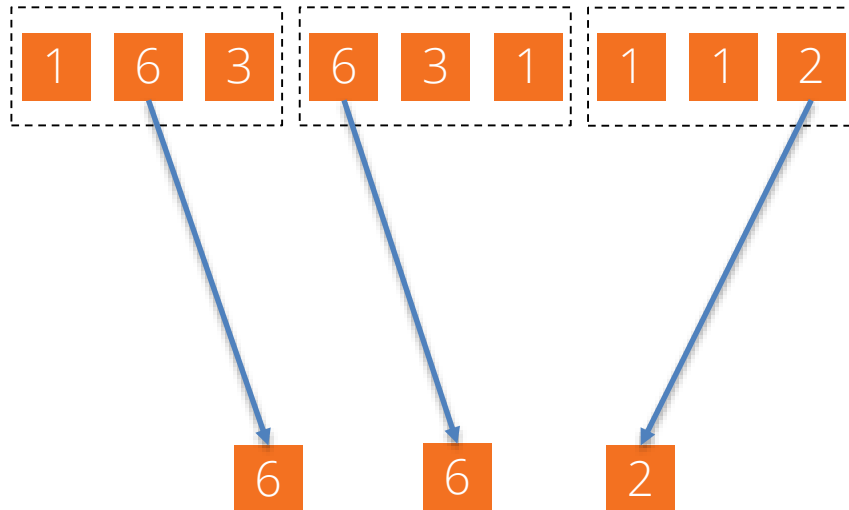


# Convolution Layer - Filters

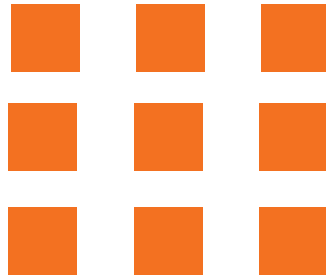


# Max Pooling

pool\_size = 3



# Flattening



# Dogs vs Cats

- Example from keras book
  - <https://www.manning.com/books/deep-learning-with-r>



dog.1.jpg



dog.2.jpg



cat.1.jpg



cat.2.jpg



dog.7.jpg



dog.8.jpg



cat.7.jpg



cat.8.jpg



dog.13.jpg



dog.14.jpg



cat.13.jpg



cat.14.jpg



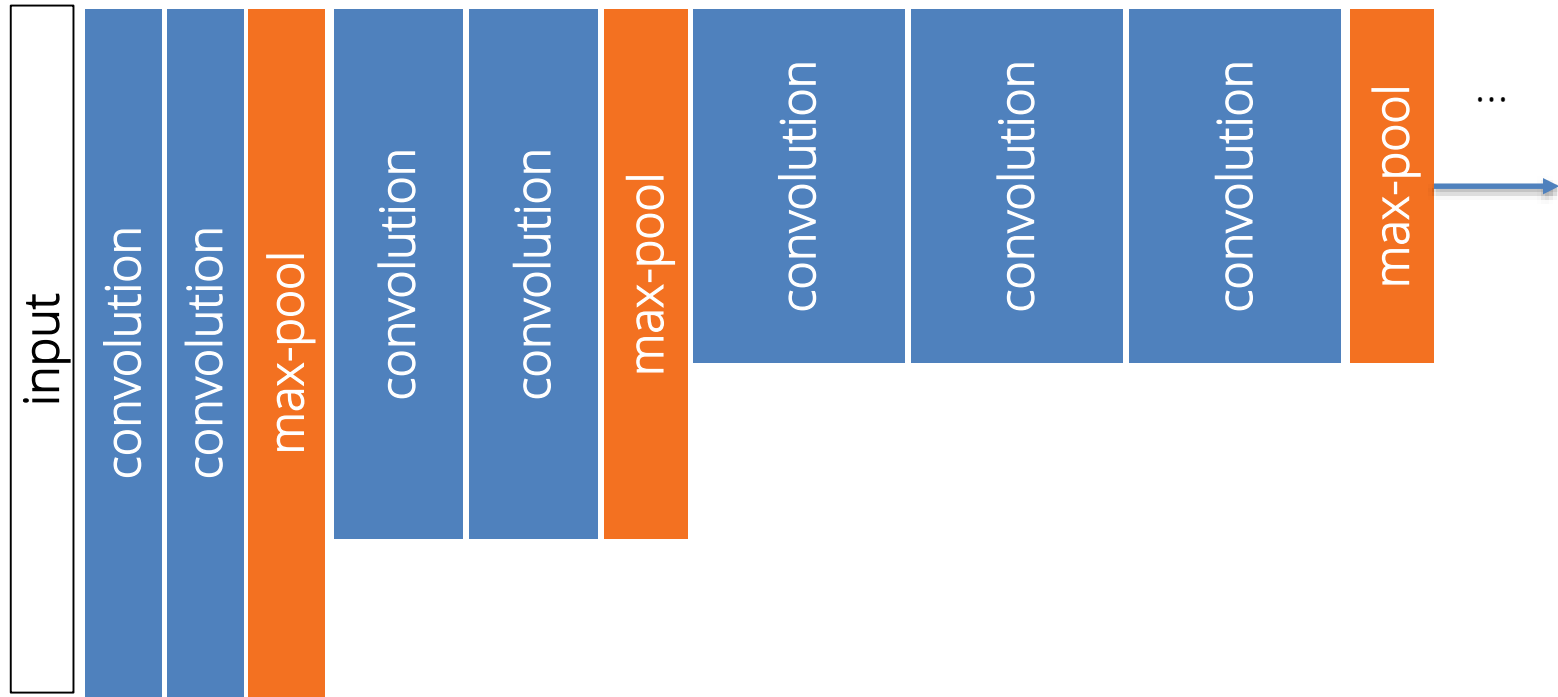
# Dogs vs Cats

- Images stored in Cloud Storage
  - Keras `flow_images_from_directory`
- Requires GPU
- 03 – Train model from scratch
- 04 – Build on pretrained-model



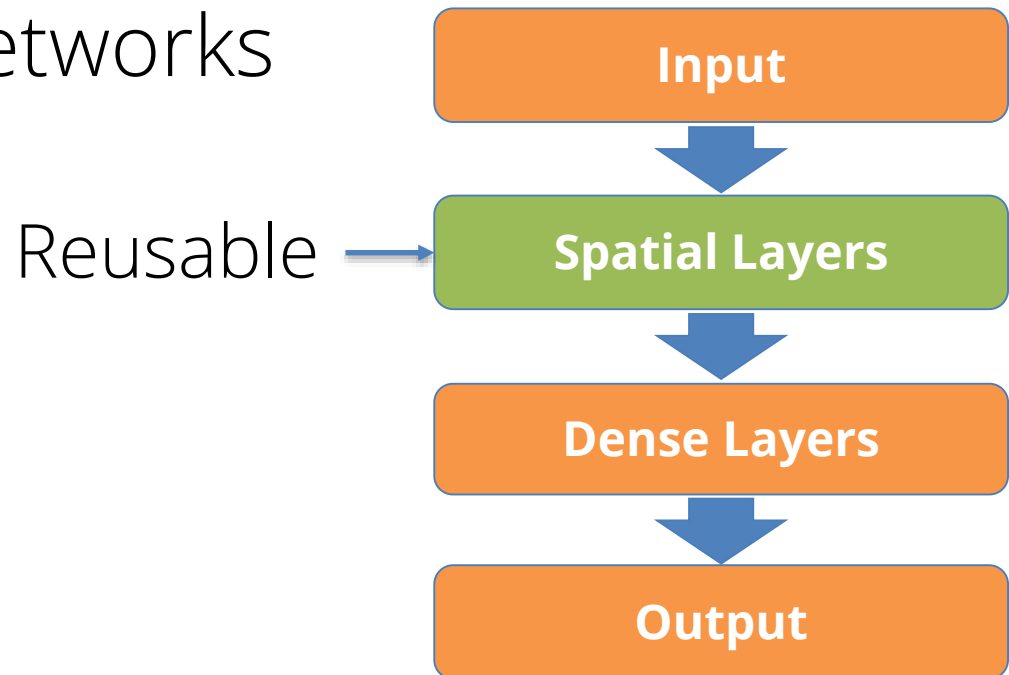


# CNN Architectures - VGG



# Pre-trained Networks

- Pre-trained Networks



# Summary

- Cloudml is easy
  - Use it before you buy a GPU
  - Use it for any compute task
- Don't feel bad using R for TensorFlow
  - Just get stuff done!
  - It's just another layer

