Google Cloud Platform for Data Science

DATA EXPLORATION, VISUALIZATION, COLLABORATION & DATA ANALYSIS

Google Cloud Platform

Sofia Yfantidou
Syeda Noor Zehra Naqvi

Big Data Lifecycle

...FROM RAW DATA TO INSIGHTS

DATA COLECTION

DATA CLEANING & PREPROCESSING

DATA VISUALIZATION & COLLABORATION

DATA EXTRACTION & TRANSFORMATION

DATA EXPLORATION

DATA ANALYSIS

BigQuery

DATA EXPLORATION

Serverless Data Warehouse.

Find insights using familiar SQL.

CONNECT TO DATASETS

QUERY DATASETS

Access

Programmatical Integration

- REST API

Direct Access

- Web UI
- bq Command Line

```
#standardSQL
SELECT
  location,
  country,
  value,
  latitude,
  longitude
FROM
  `bigquery-public-data.openaq.global_air_quality`
WHERE
  pollutant = "pm10"
```

Cloud Datalab

DATA EXPLORATION, COLLABORATION AND VISUALIZATION

Interactive tool built on Jupyter.

Explore, visualize and analyse data.

Google Cloud ML Engine

BigQuery

Cloud Storage

Jupyter Notebook

Python

Standard SQL

Cloud Datalab

DATA EXPLORATION, COLLABORATION AND VISUALIZATION

Interactive notebooks to code and collaborate

Instantiate Google Compute Engine

Google Datalab

DATA EXPLORATION, COLLABORATION AND VISUALIZATION

Integrated and Open-source

Multi-language Support

Notebook Format and IPython Support

Machine Learning

Google Data Studio

DATA VISUALIZATION AND COLLABORATION

Tool to visualize the data sets and create extensive reports/dashboards.

CONNECT and ADJUST DATASETS

CREATE REPORTS/DASHBOARDS

Google Data Studio

DATA VISUALIZATION AND COLLABORATION

Free Beta Version

15 Google Connectors and 74 Partner Connectors

Simple Drag and Drop Interface

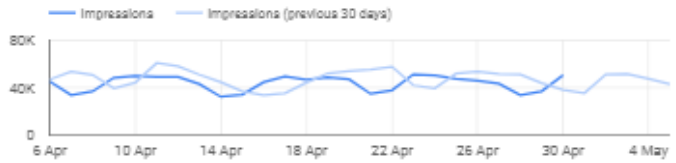
Share and Collaborate

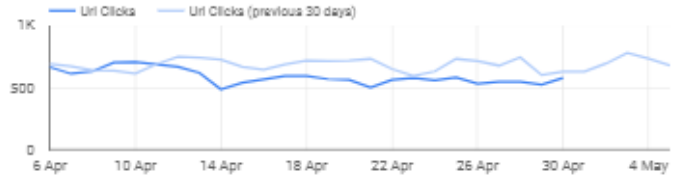
Google Search Console

6 Apr 2018 - 5 May 2018

Filters | Country | Device Category

	Impressions	Url Clicks	URL CTR
Last 28 days	1.1M	14.6K	1.3%
Previous 28 days	▼ -23.2%	▼ -28.4%	▼ -6.7%





Landing Page	Impressions	Url Clicks	URL ...
https://shop.googlemerchandisestore.com/Google+R...			0.06%
https://shop.googlemerchandisestore.com/Google+R...			1.11%
https://shop.googlemerchandisestore.com/Google+R...			3.1%
https://shop.googlemerchandisestore.com/Google+R...			1.93%
https://shop.googlemerchandisestore.com/Google+R...			0.38%
https://shop.googlemerchandisestore.com/Google+R...			0.5%
https://shop.googlemerchandisestore.com/			3.35%
https://shop.googlemerchandisestore.com/store-poli...			1.06%
https://shop.googlemerchandisestore.com/Google+R...			0.98%
https://shop.googlemerchandisestore.com/Google+R...			1.84%
https://shop.googlemerchandisestore.com/Google+R...			0.5%
https://shop.googlemerchandisestore.com/Google+R...			0.35%

Query	Impressions	Clicks	Site ...	Average Position
waze			0.01%	8.82
google t			0.01%	5.12
			3.96%	15.48
youtube			0.07%	17.58
google shop			0.62%	7.15
google store			0.2%	9.9
apparel			0.08%	12.07
merchandise			0.1%	10.31
google youtube			0.19%	9.45
google water			0.1%	6.23
youtube google			0.29%	10.1
google kids			0.07%	9.26

Google Machine Learning Engine

DATA ANALYSIS

Build superior models

Deploy them into production

TRAINING SERVICES

PREDICTION SERVICES

Google Machine Learning Engine

DATA ANALYSIS

TRAINING

Google Machine Learning Engine

DATA ANALYSIS

PREDICTION

ONLINE PREDICTION

Real-Time Applications

BATCH PREDICTION

Asynchronous Applications

Google Machine Learning Engine

DATA ANALYSIS

Automatic Resource Provisioning and Scalability

Automatic Hyper-tuning

Integrated Services

Model Portability

TensorFlow

DATA ANALYSIS

Open-source machine learning library for high performance numerical computation.

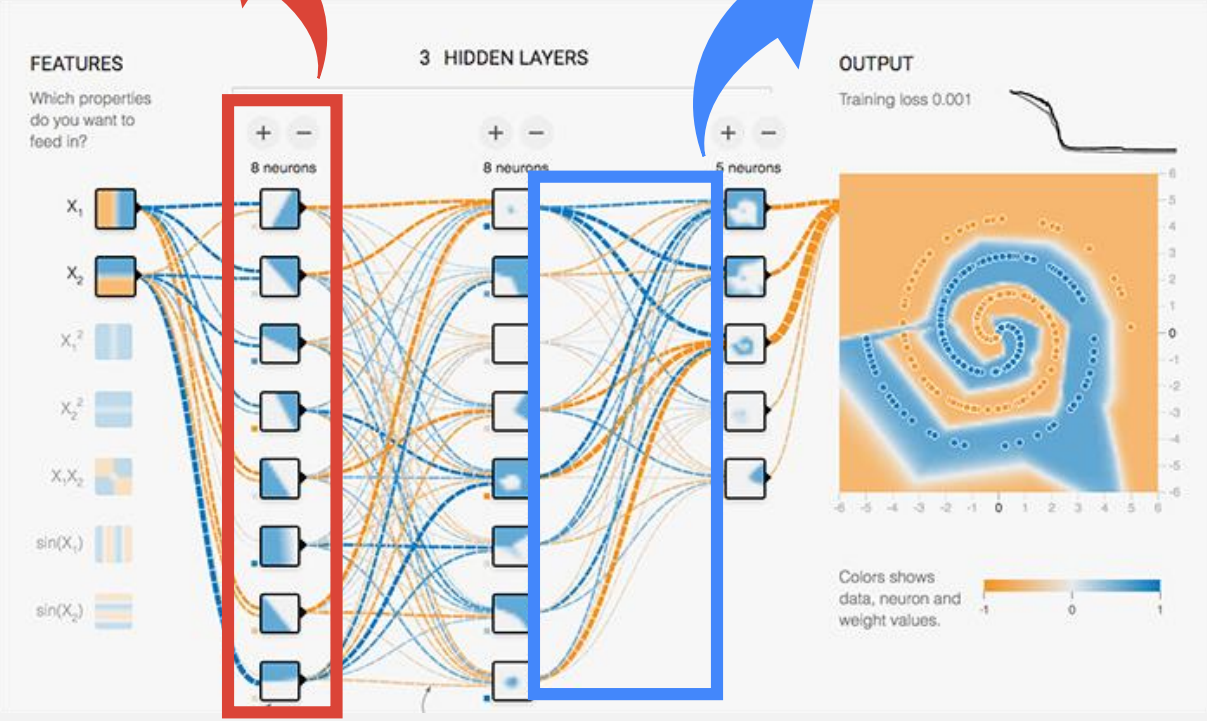
TensorFlow

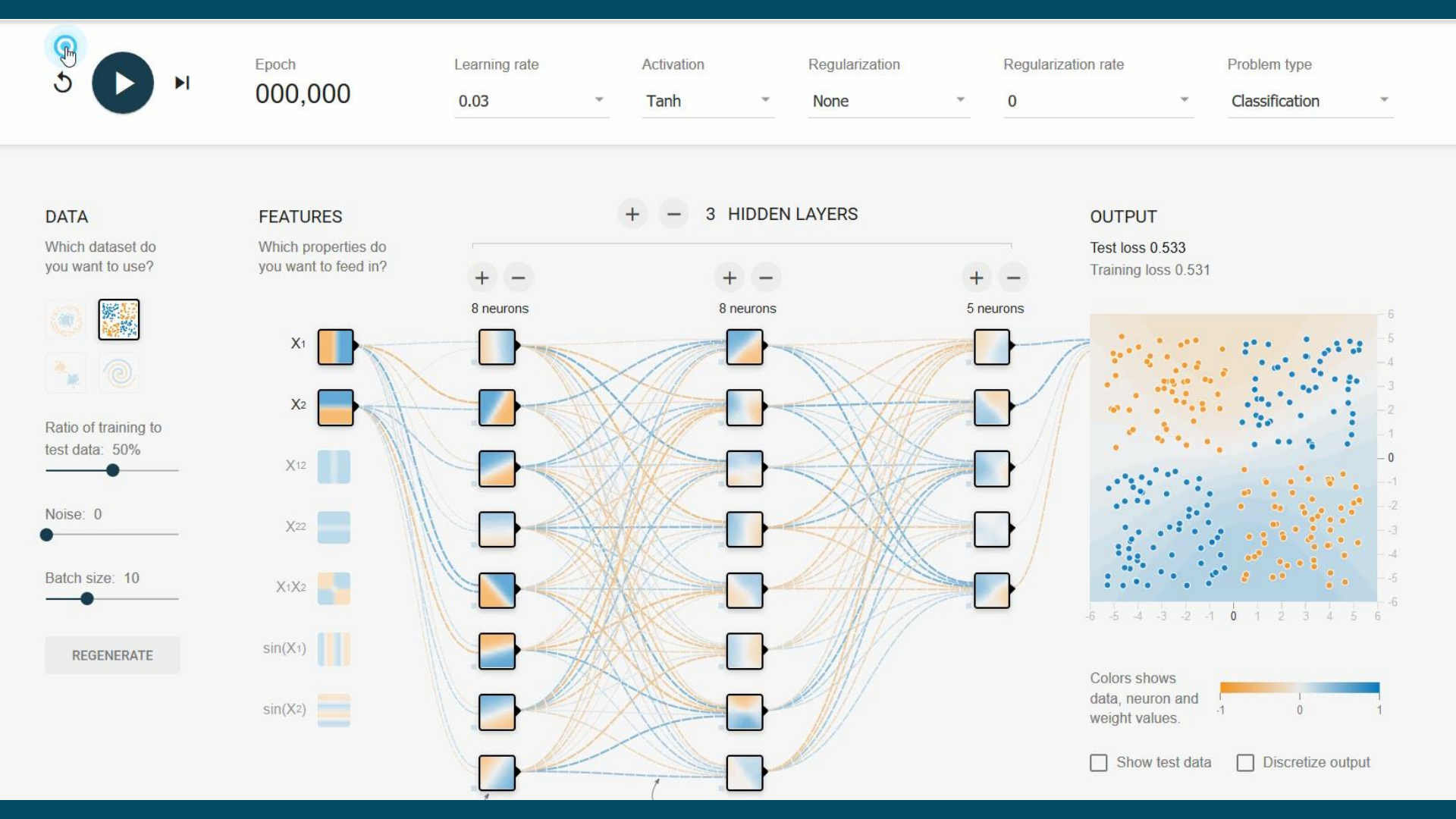
Nodes = Mathematical Operations

DATA ANALYSIS

Edges = Data Produced or Consumed

Dataflow Graph




Epoch
000,000
Learning rate
0.03
Activation
Tanh
Regularization
None
Regularization rate
0
Problem type
Classification
DATA
Which dataset do you want to use?
Ratio of training to test data: 50%
Noise: 0
Batch size: 10
REGENERATE
FEATURES
Which properties do you want to feed in?
X1
X2
X12
X22
X1X2
sin(X1)
sin(X2)
3 HIDDEN LAYERS
8 neurons
8 neurons
5 neurons
OUTPUT
Test loss 0.533
Training loss 0.531
Colors shows data, neuron and weight values.
-1
0
1
Show test data
Discretize output

TensorFlow

DATA ANALYSIS

Features

Power

Adaptability

Flexibility

Portability

THANKS!

Google Cloud Platform

For details: https://github.com/syfantid/Research-topic