toptal

Kickstart your career in ML

/bojorges

@eeveeta

/eva-bojorges

About me…

MSc Cognitive Science

Machine Learning Engineer:
Health, Security,
Environmental Science

Toptal Latin America &
Africa Regional Leader

MMA Fighter, loves cats

toptal

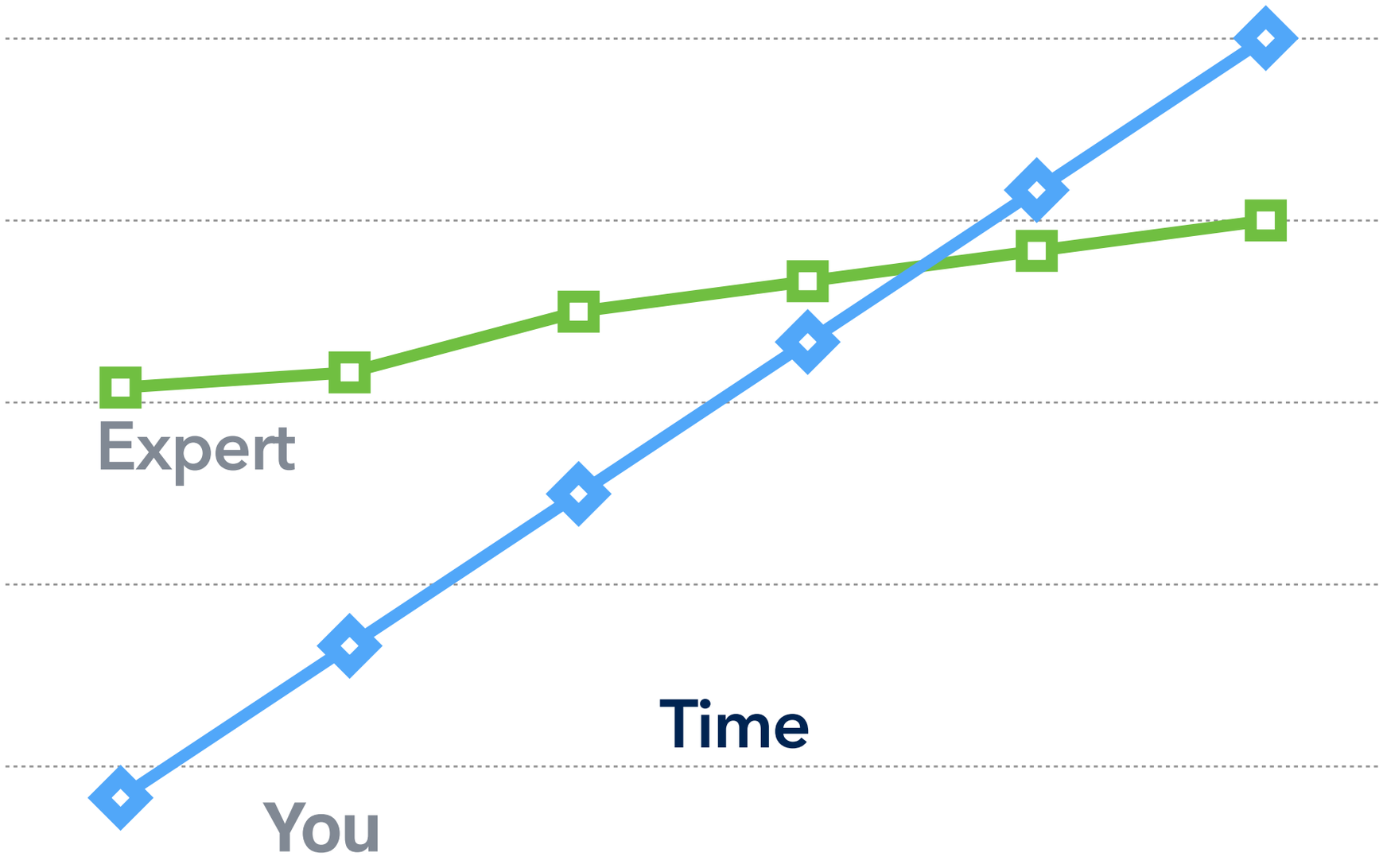
@eeveeta

/bojorges

www.toptal.com/medellin-ml

$$\delta w_{ji} = \alpha\left(t_j - \phi\left(h_j\right)\right)\phi'\left(h_j\right)x_i$$

Learning



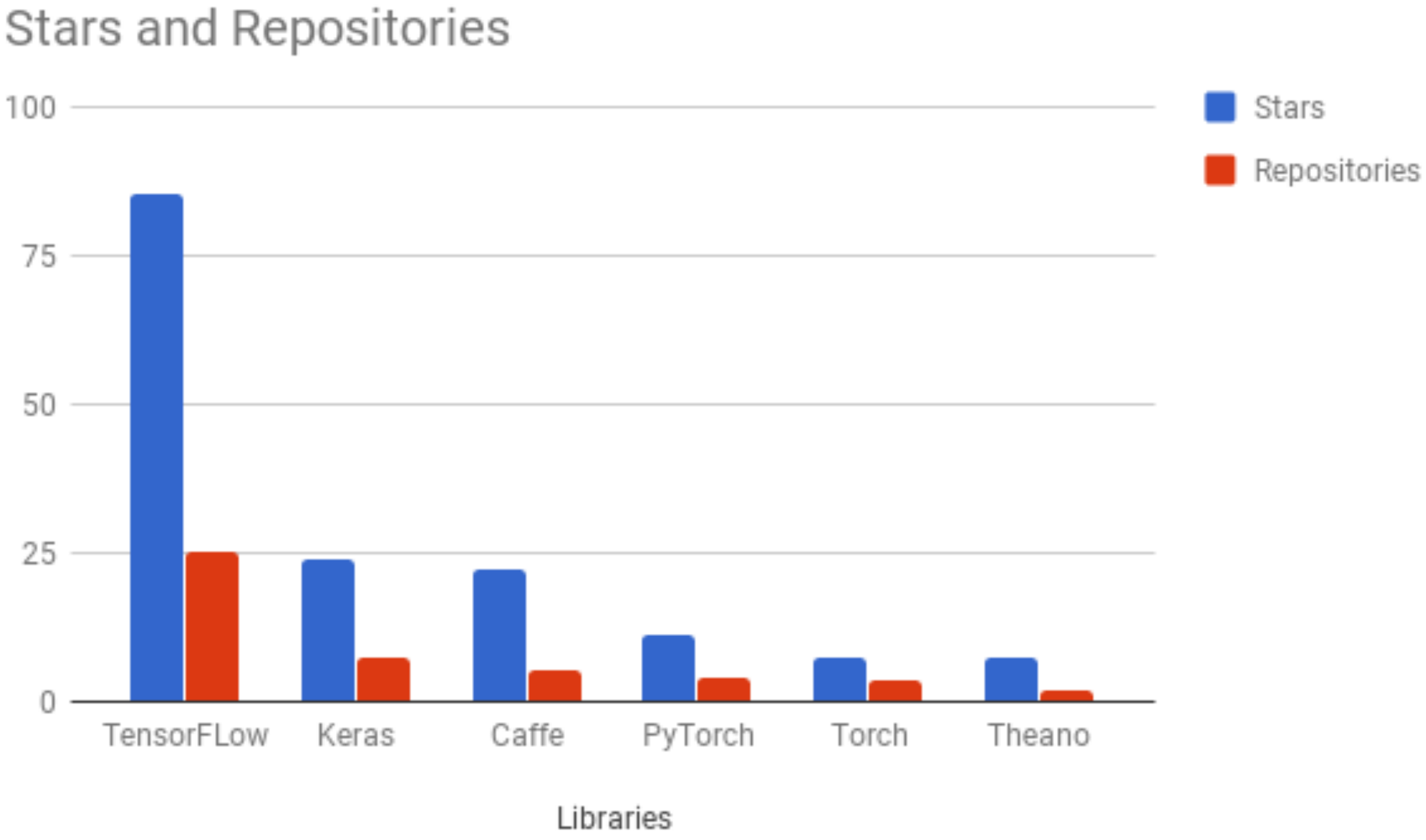
Why TensorFlow?

Denny Britz @dennybritz · 25 Dec 2017

I'm going through my newsletters to write up a year-end summary of developments and achievements in AI.

Fun fact: Almost every week, a company released a new generic or task-specific Deep Learning "framework" 😅

Why TensorFlow?



Companies using TensorFlow

Google

OpenAI

AIRBUS

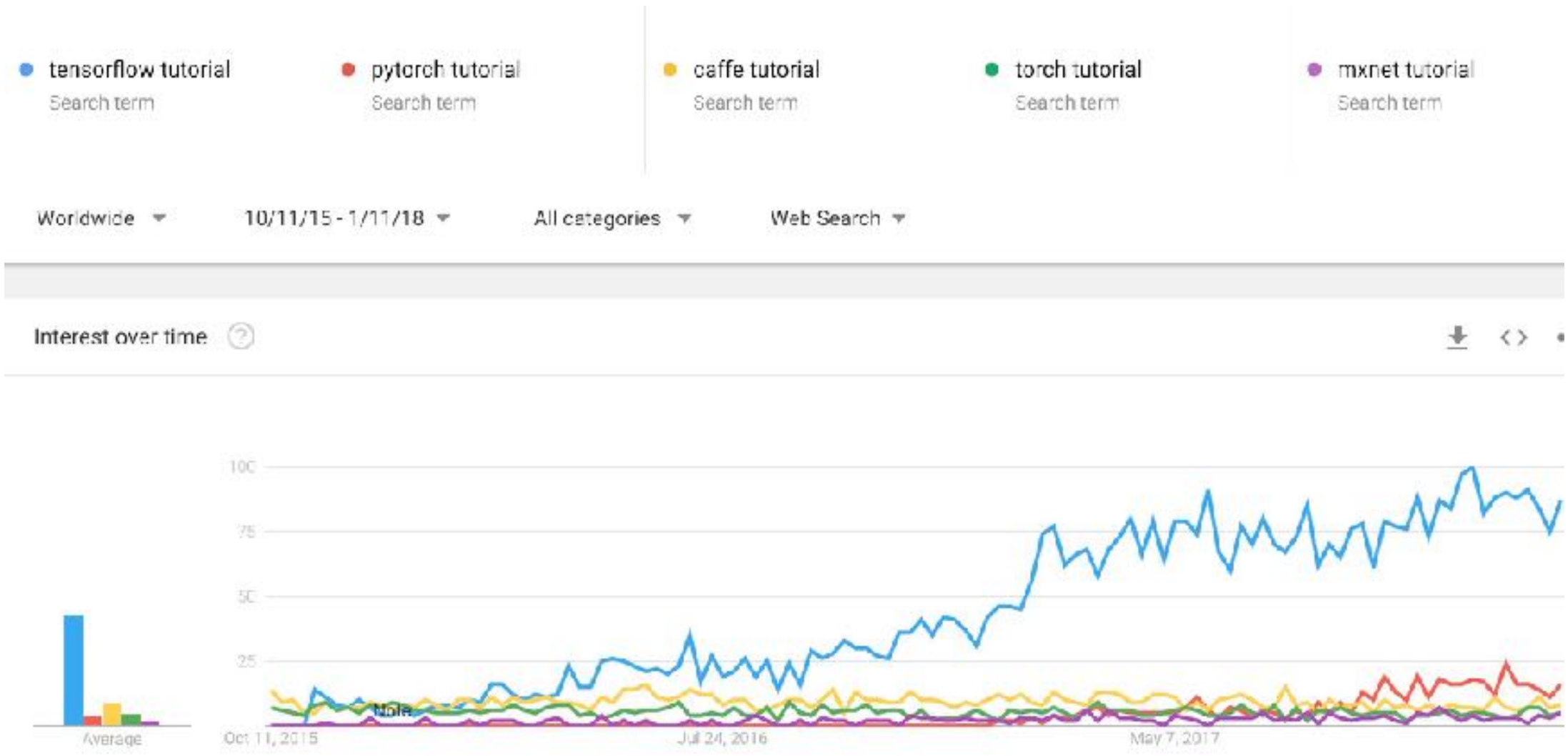
Google DeepMind

intel

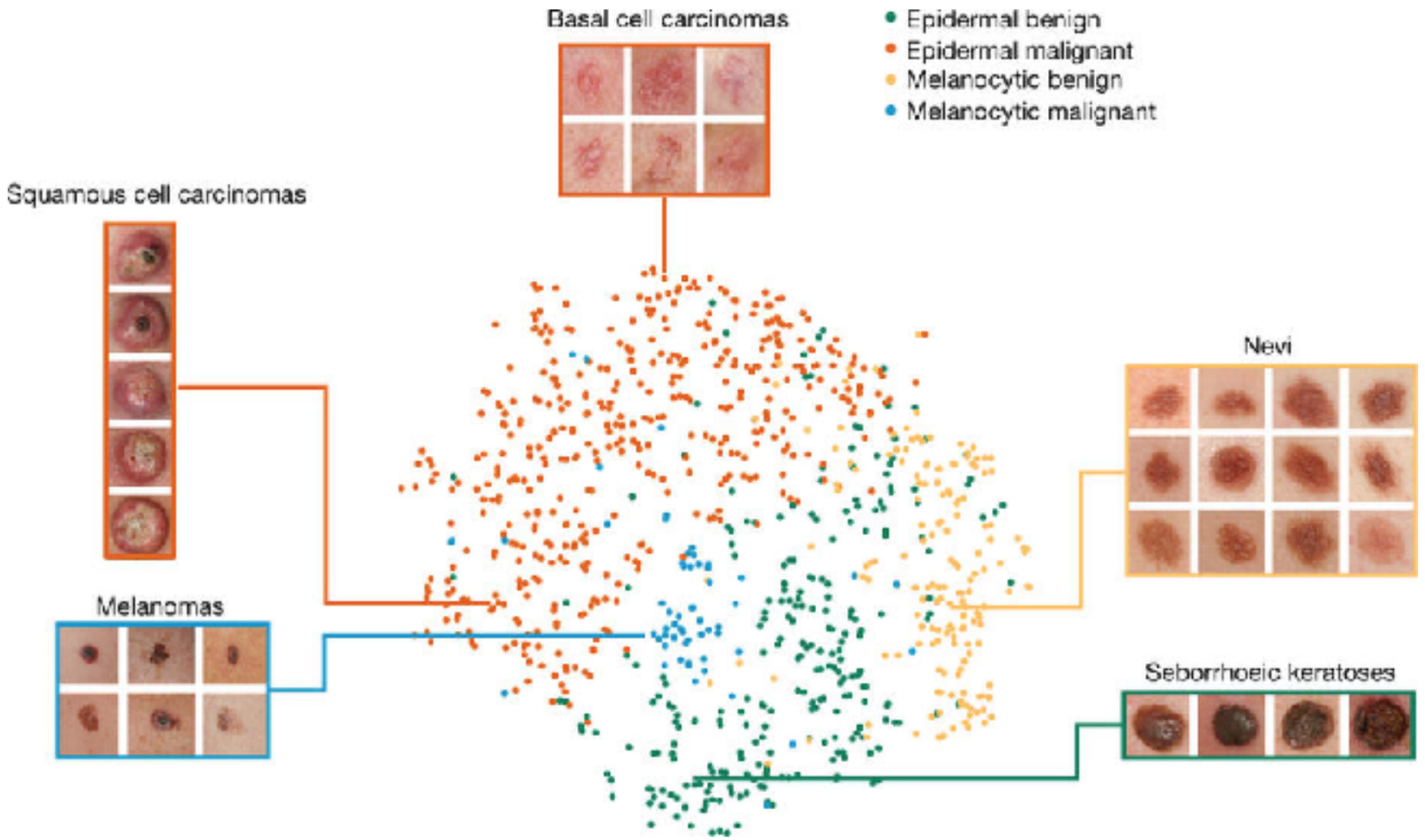
airbnb

NVIDIA

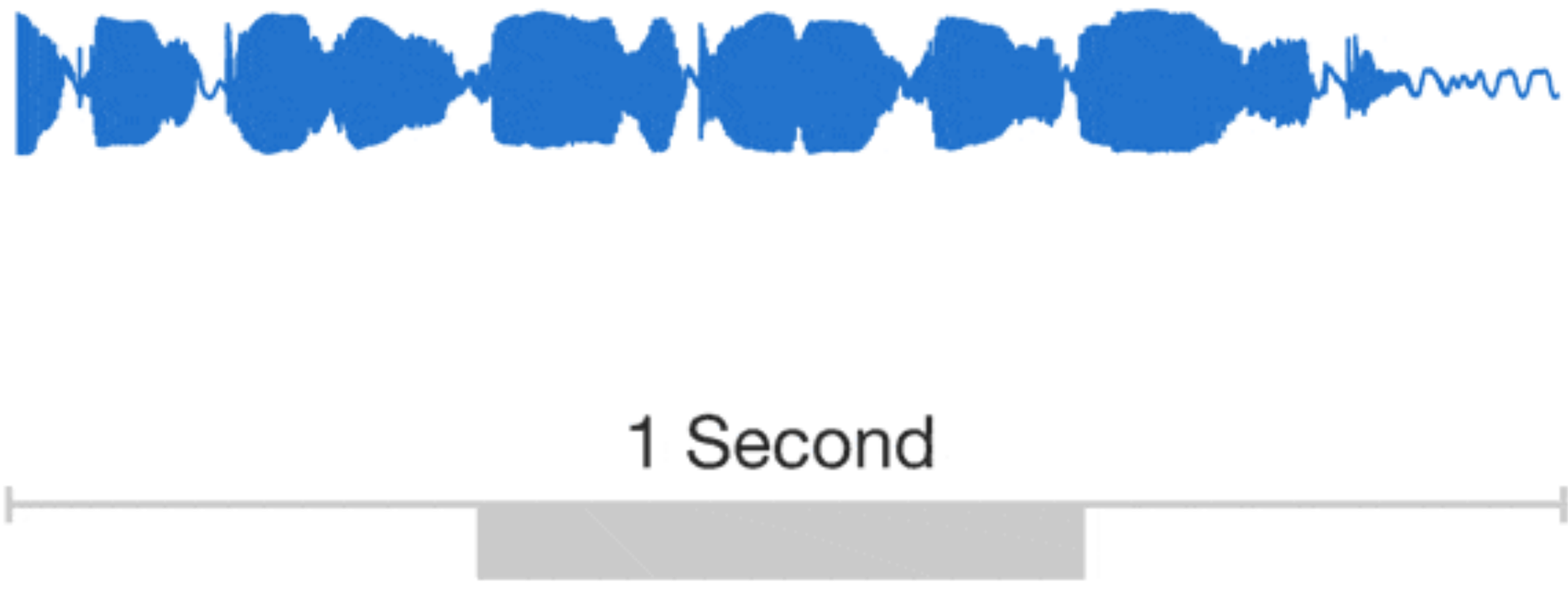
Demand for tutorials on TensorFlow



Classify skin cancer



WaveNet: Text to Speech



Drawing

Neural Style Translation

The best time to learn ~~Machine Learning~~ Deep Learning.

Tensorflow launched November 2015

Andrew Ng's
Deep Learning Specialization launched
August 2017

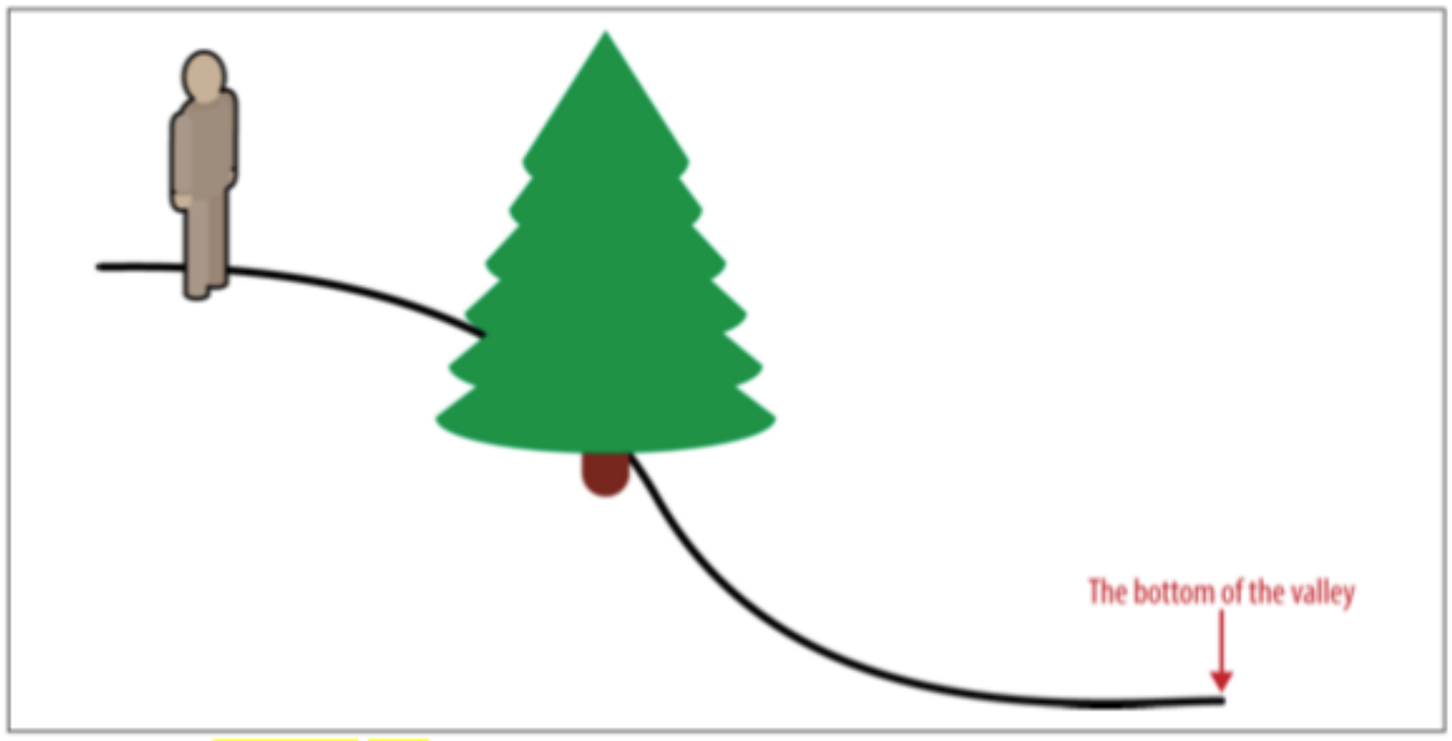
Sequence models available since February 2018

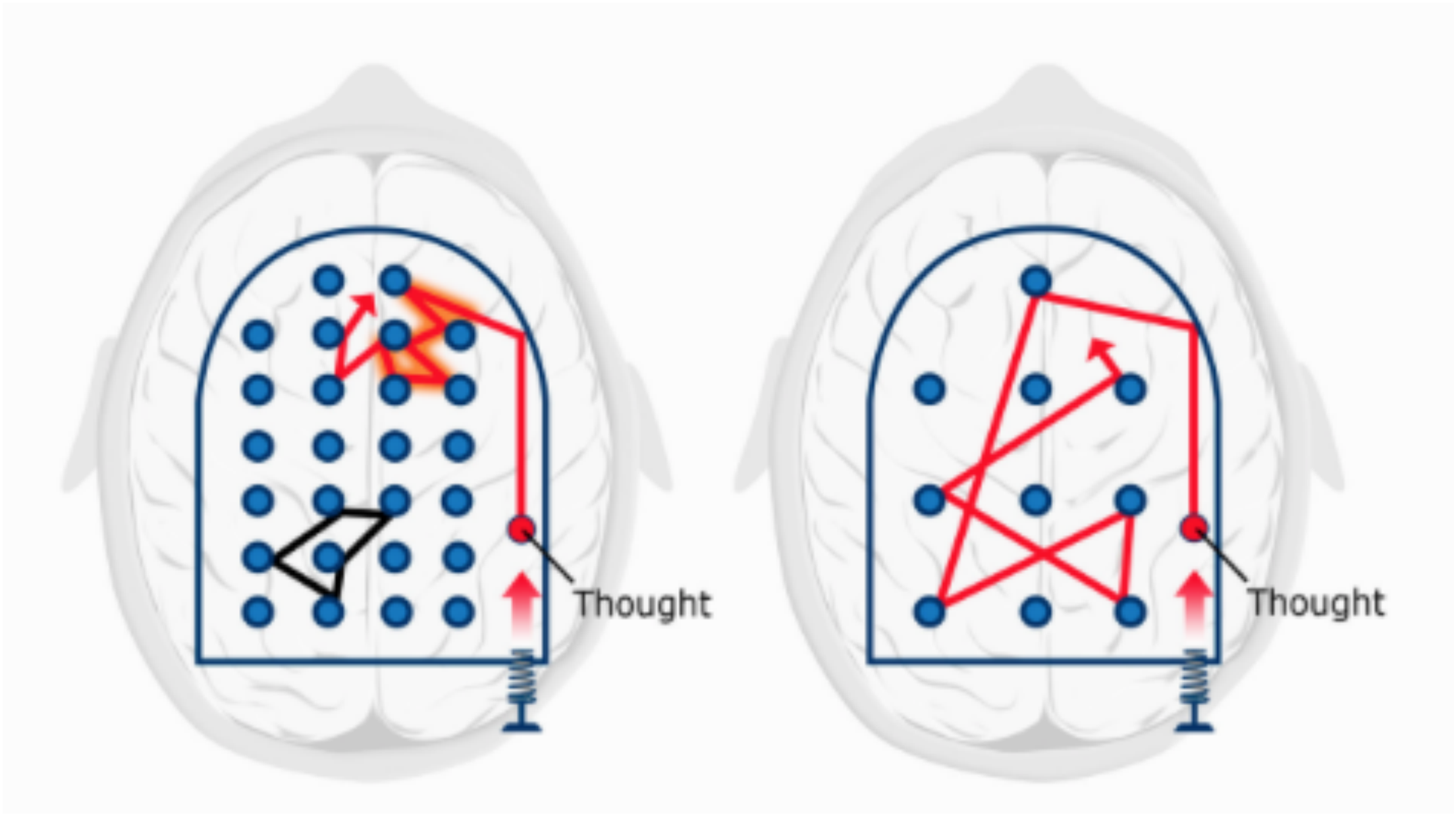
Learning Tensorflow

Gradient Descent

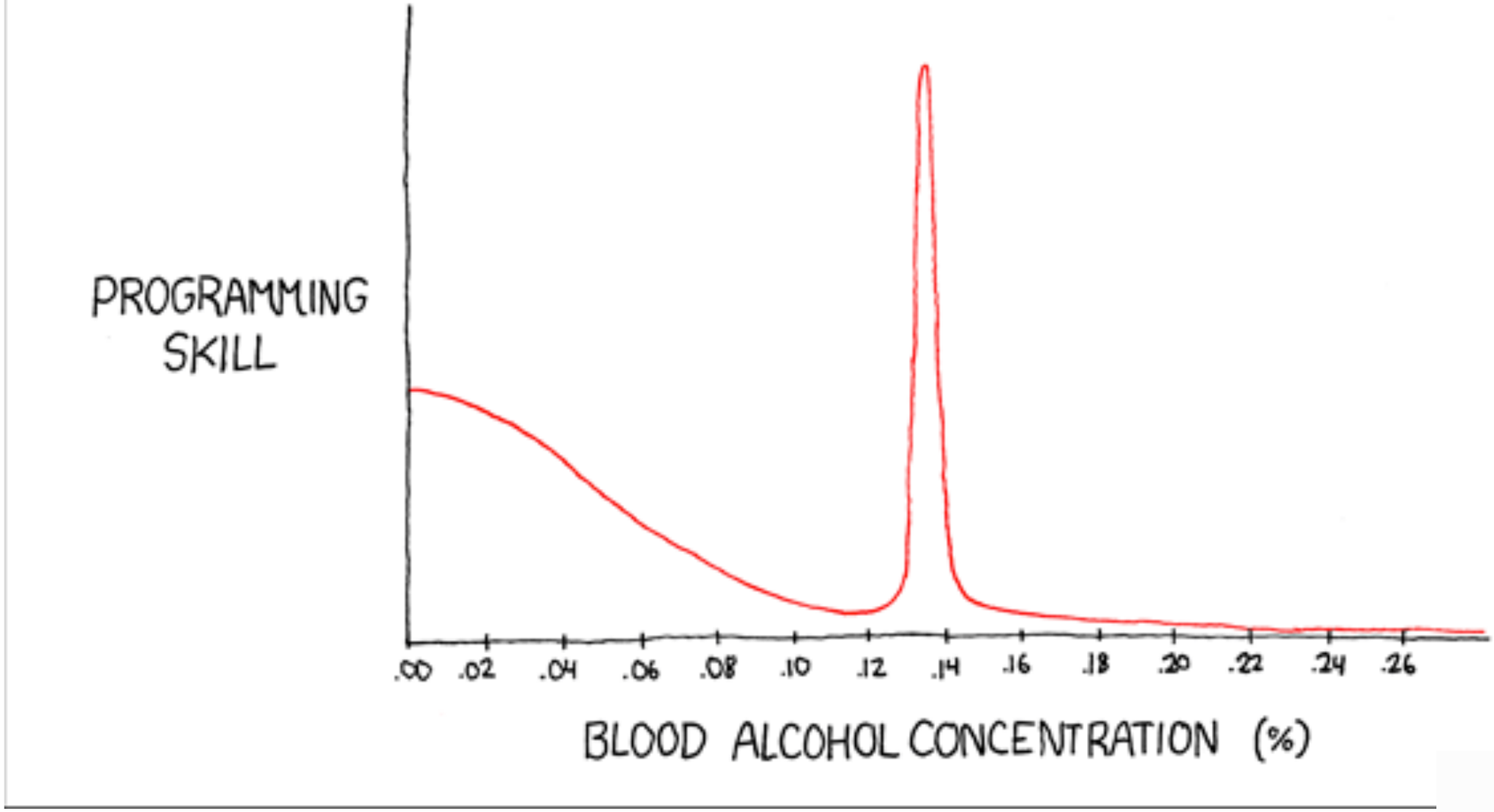
$$\delta w_{ji} = \alpha\left(t_j - \phi\left(h_j\right)\right)\phi'\left(h_j\right)x_i$$

```
alpha * (expected - calculated) * derivative_of_calculated * input_at_i
```



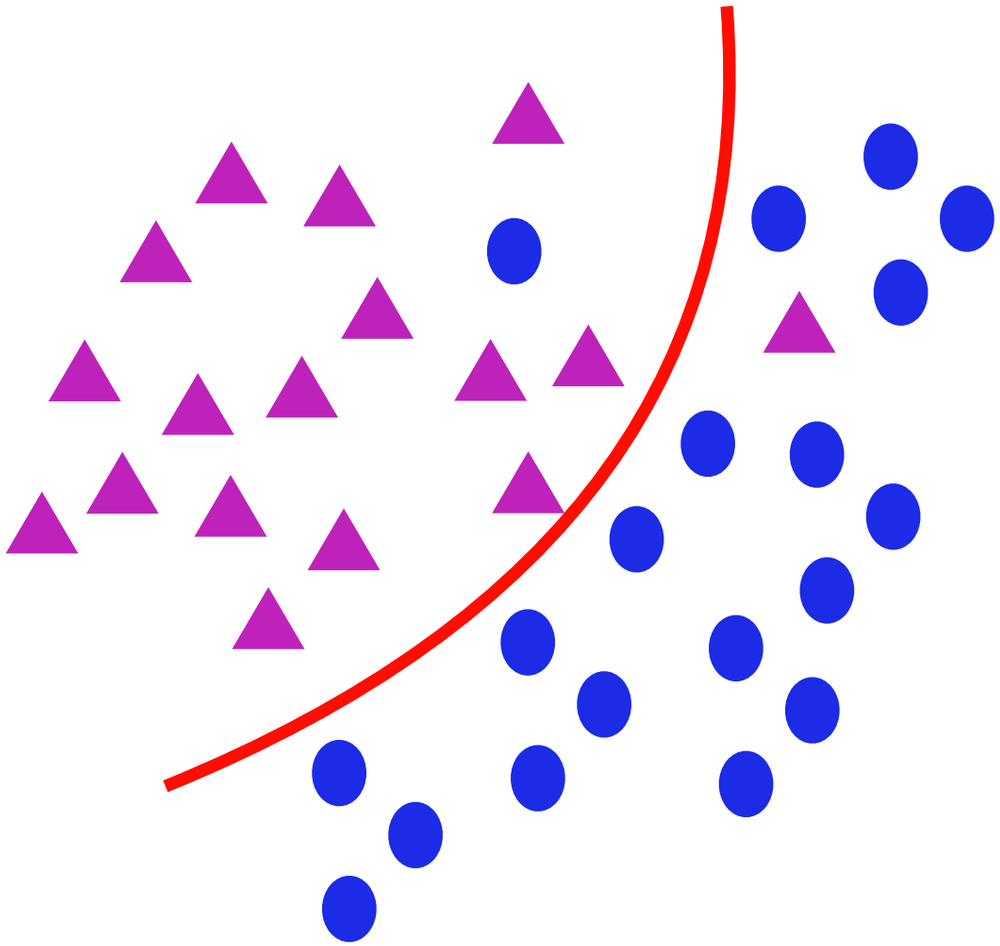

Thought
Thought

Ballmer Peak

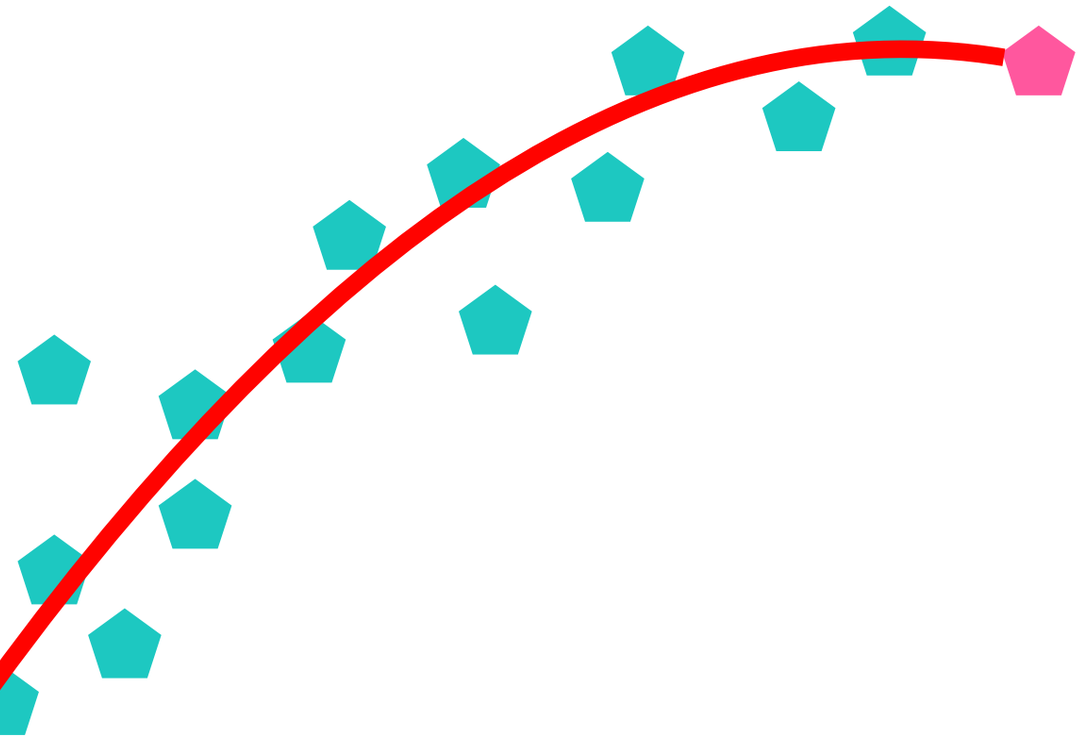


X			y
Time of the day	Language of choice	Bottles of beer	# of bugs
13:00	PHP	0	16
15:00	Ruby	2	8
20:00	Python	1	7
23:00	JavaScript	5	9
3:00	Haskell	I forgot	4

MACHINE LEARNING METHODS

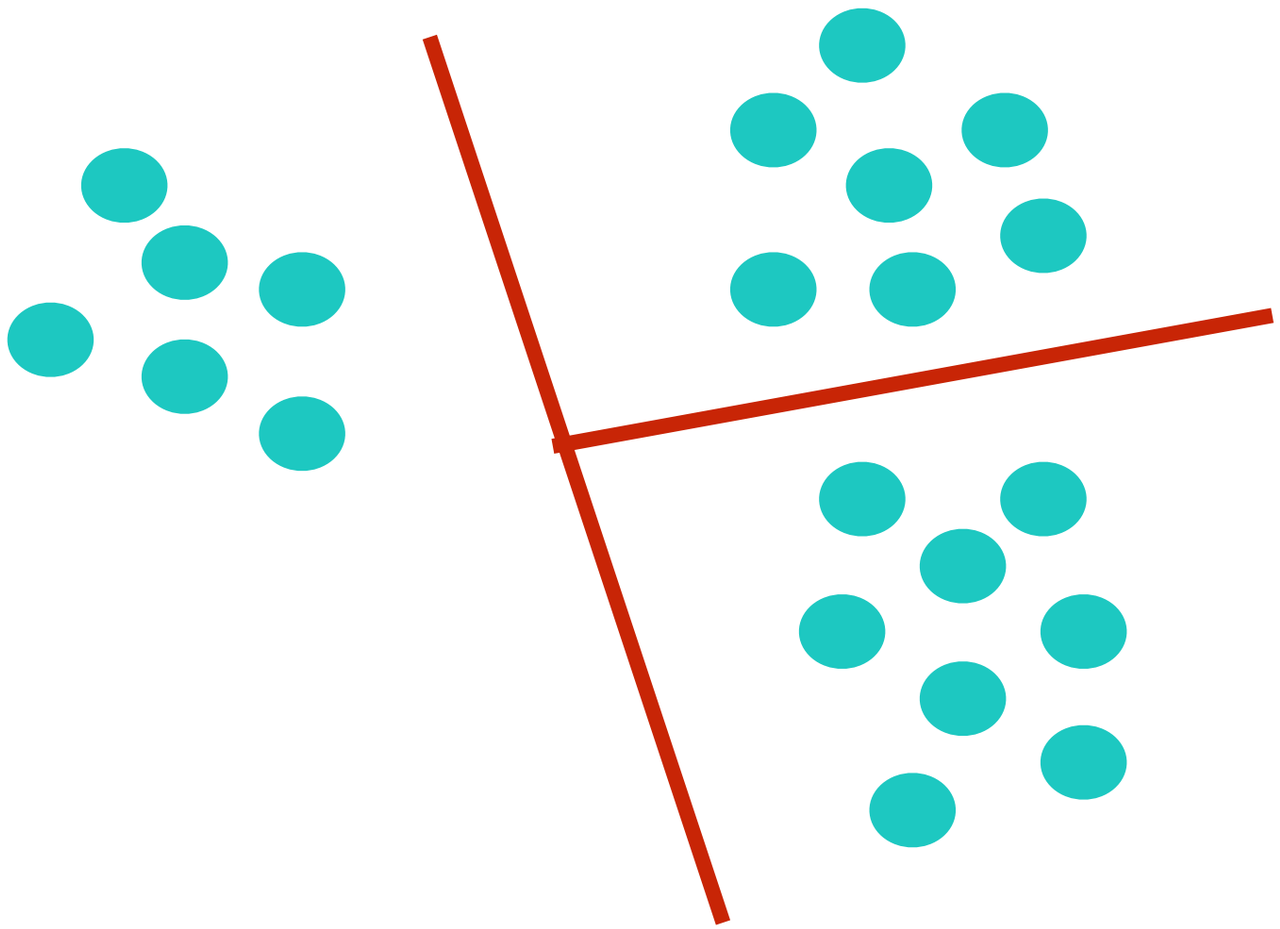


CLASSIFICATION



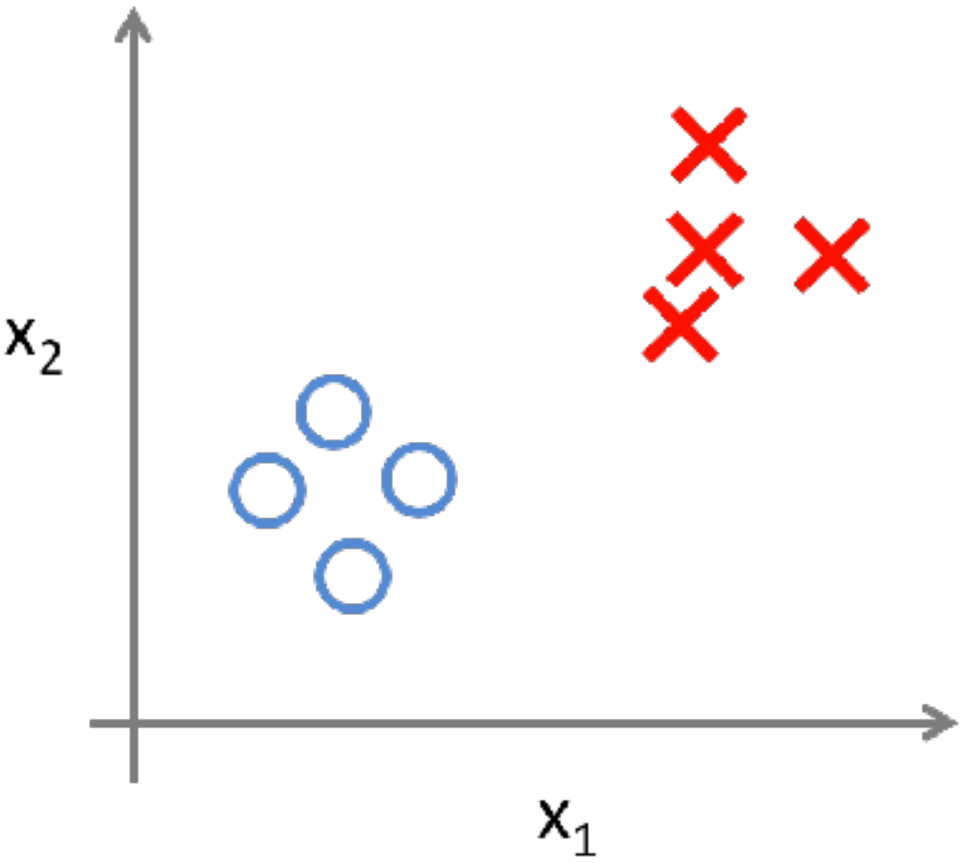
REGRESSION

MACHINE LEARNING METHODS

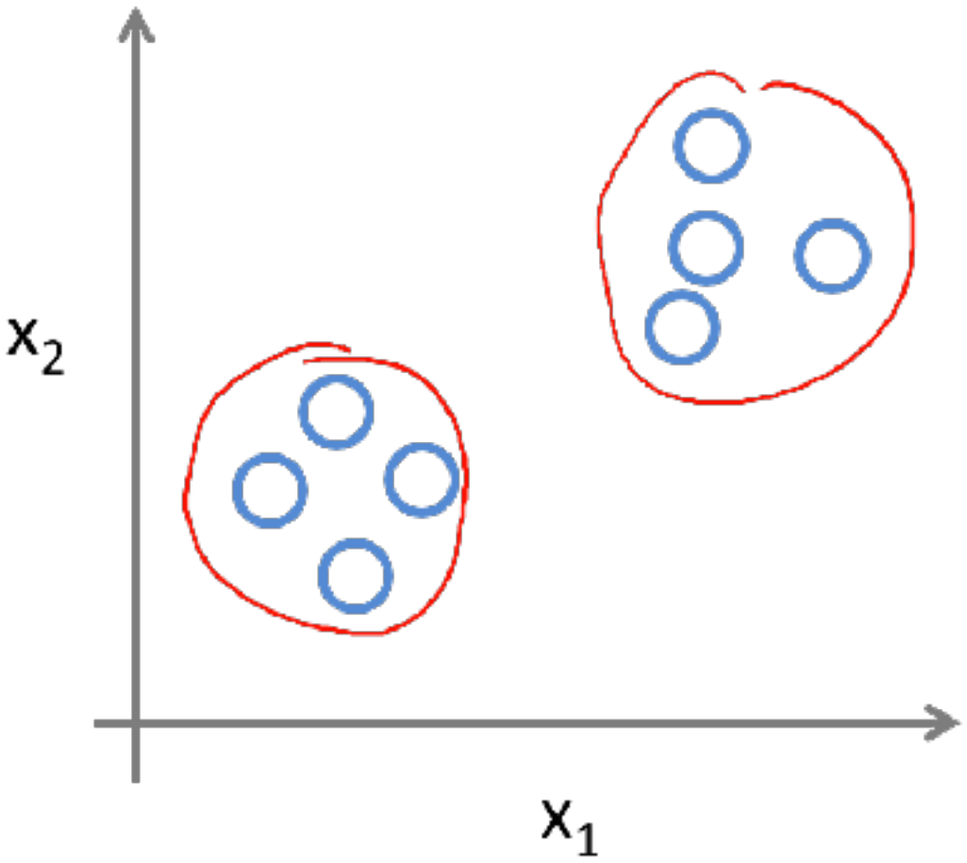


CLUSTERING

Supervised Learning



Unsupervised Learning



X			y
Time of the day	Language of choice	Bottles of beer	# of bugs
13:00	PHP	0	16
15:00	Ruby	2	8
20:00	Python	1	7
23:00	JavaScript	5	9
3:00	Haskell	-	4

Fill up this value

```
1 from sklearn.preprocessing import Imputer
2
3 imputer = Imputer(strategy="median")
4
5 code_num = code_data.drop("language", axis=1)
6
7 imputer.fit(code_num)
8
9 X = imputer.transform(code_num)
10
```

Min-max scaling	**Standarization**
0-1	0 mean, unit variance
Values shifted and rescaled	Good with outliers
MinMaxScaler	StandardScaler

X			y
Time of the day	Language of choice	Bottles of beer	# of bugs
13:00	PHP	0	16
15:00	Ruby	2	8
20:00	Python	1	7
23:00	JavaScript	5	9
3:00	Haskell	-	4

Label Encoding

Name	Number
PHP	1
Ruby	2
Python	3
JavaScript	4
Haskell	5

Similarity

Haskell

JavaScript

Python

Ruby

PHP

Similarity

JavaScript

PHP

Python

Ruby

Haskell

One Hot Encoding

Time of the day	Bottles of beer	# of bugs	PHP?	Ruby?	Python?
13:00	0	16	1	0	0
15:00	2	8	0	1	0
20:00	1	7	0	0	1
23:00	5	9	0	0	0
3:00	-	4	0	0	0

```
from sklearn.pipeline import FeatureUnion

num_pipeline = Pipeline([
        ('selector', DataFrameSelector(num_attribs)),
        ('imputer', Imputer(strategy="median")),
        ('attribs_adder', CombinedAttributesAdder()),
        ('std_scaler', StandardScaler()),
    ])

cat_pipeline = Pipeline([
        ('selector', DataFrameSelector(cat_attribs)),
        ('label_binarizer', LabelBinarizer()),
    ])

full_pipeline = FeatureUnion(transformer_list=[
        ("num_pipeline", num_pipeline),
        ("cat_pipeline", cat_pipeline),
    ])
```

Aurélien Géron. “Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.”

Try some models

```
names = ["Nearest Neighbors", "Linear SVM", "RBF SVM", "Gaussian Process",
         "Decision Tree", "Random Forest", "Neural Net", "AdaBoost",
         "Naive Bayes", "QDA"]

classifiers = [
    KNeighborsClassifier(3),
    SVC(kernel="linear", C=0.025),
    SVC(gamma=2, C=1),
    GaussianProcessClassifier(1.0 * RBF(1.0), warm_start=True),
    DecisionTreeClassifier(max_depth=5),
    RandomForestClassifier(max_depth=5, n_estimators=10, max_features=1),
    MLPClassifier(alpha=1),
    AdaBoostClassifier(),
    GaussianNB(),
    QuadraticDiscriminantAnalysis()]
```

Iterate

```
# iterate over datasets
for ds_cnt, ds in enumerate(datasets):
    # preprocess dataset, split into training and test part
    X, y = ds
    X = StandardScaler().fit_transform(X)
    X_train, X_test, y_train, y_test = \
        train_test_split(X, y, test_size=.4, random_state=42)

    # iterate over classifiers
    for name, clf in zip(names, classifiers):
        clf.fit(X_train, y_train)
        score = clf.score(X_test, y_test)
```

http://scikit-learn.org/stable/

Classification

Identifying to which category an object belongs to.

Applications: Spam detection, Image recognition.
Algorithms: SVM, nearest neighbors, random forest, ... — Examples

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.
Algorithms: SVR, ridge regression, Lasso, ... — Examples

Dimensionality reduction

Reducing the number of random variables to consider.

Applications: Visualization, Increased efficiency
Algorithms: PCA, feature selection, non-negative matrix factorization. — Examples

Model selection

Comparing, validating and choosing parameters and models.

Goal: Improved accuracy via parameter tuning
Modules: grid search, cross validation, metrics. — Examples

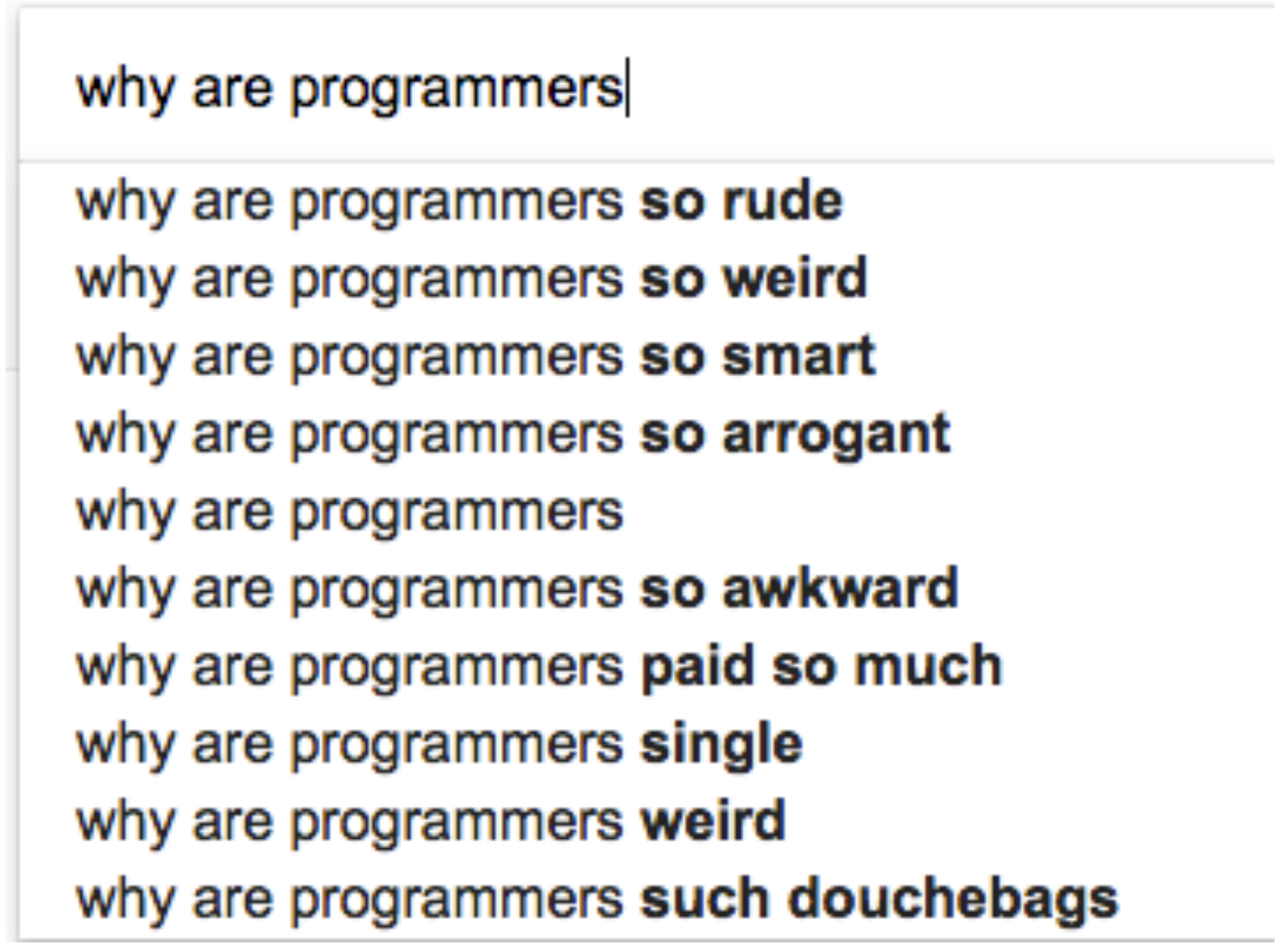
Be the worst

“Always be the worst guy in every band you’re in.”

-Pat Metheny

Get a job that challenges you.

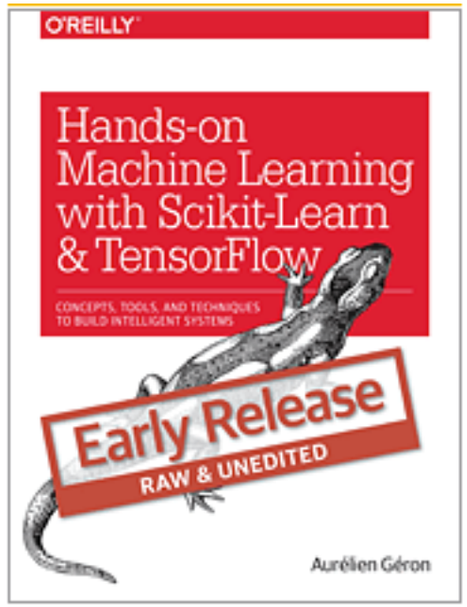
Breaking into the field



Get a job at Google

BEGINNERS

kaggle



Hands-on Machine Learning with Scikit-Learn & TensorFlow Aurélien Géron

Welch Labs Youtube

Learning