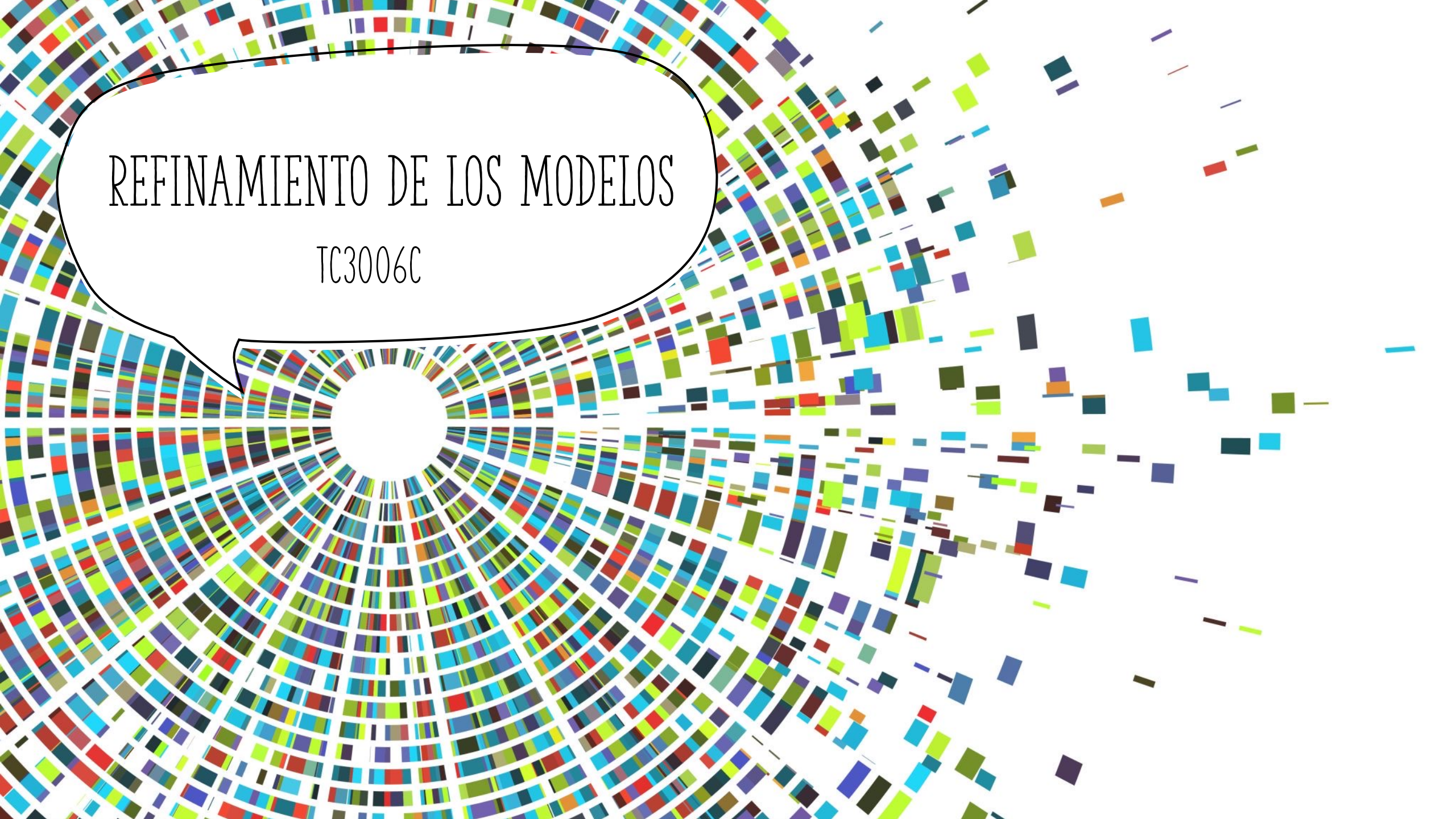
REFINAMIENTO DE LOS MODELOS

TC3006C



WHAT ARE HYPERPARAMETERS?

Parameters

Elements (variables) that a model is searching for in order to find a good fit

Linear and Logistic regressions

- m and b

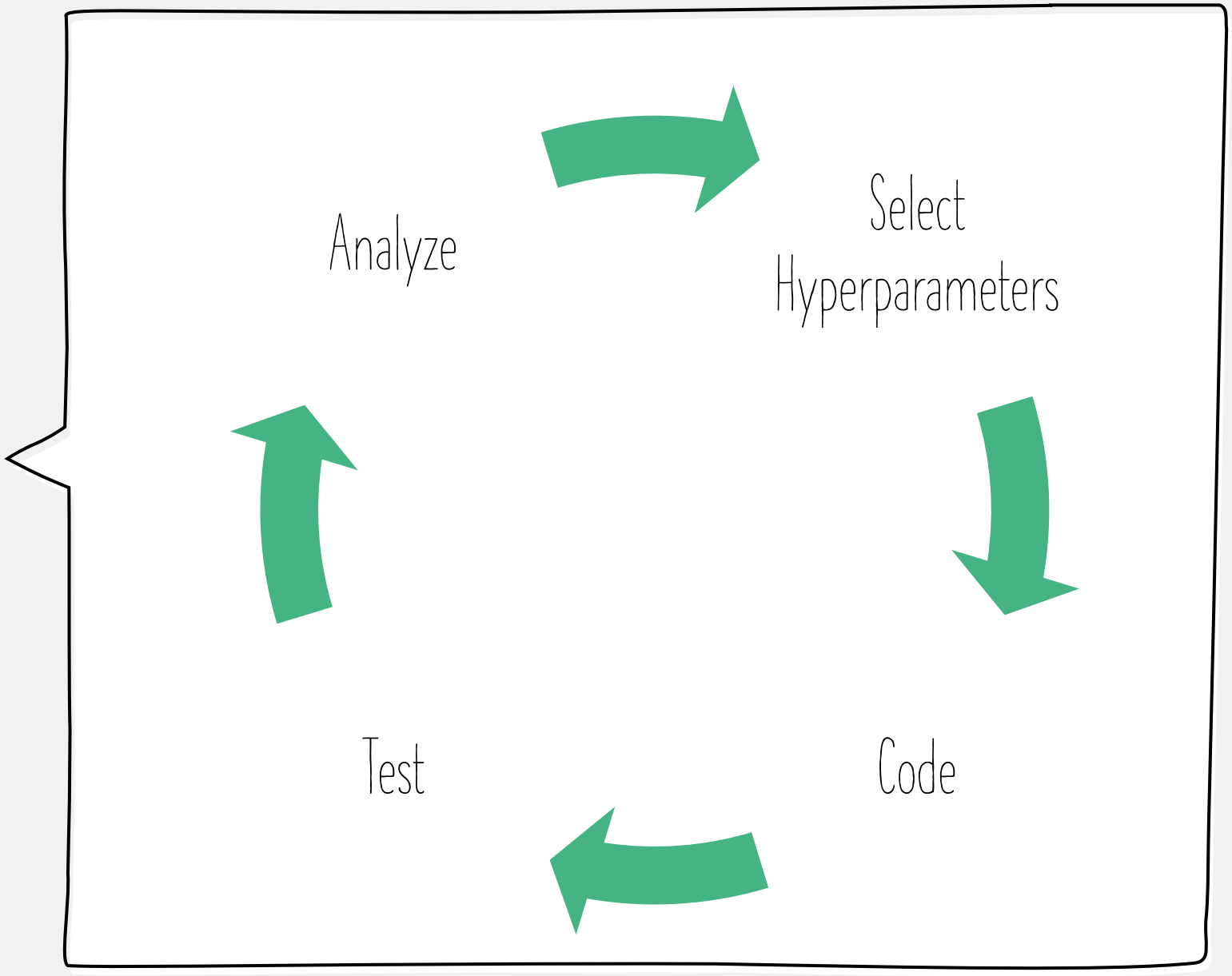
Hyperparameters

Elements (variables) that control the performance of a learning algorithm

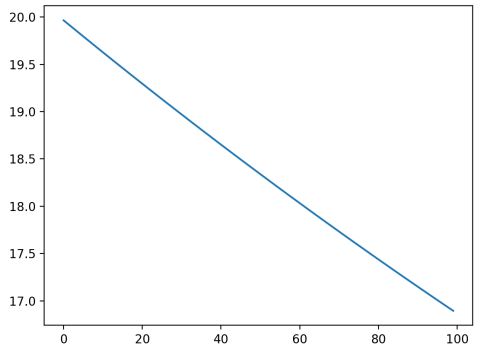
Linear and Logistic regressions

- α (learning rate)
- Scaling
- Batch size
- Epochs

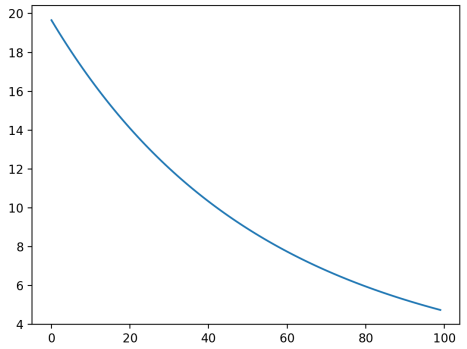
MACHINE LEARNING IS AN ART



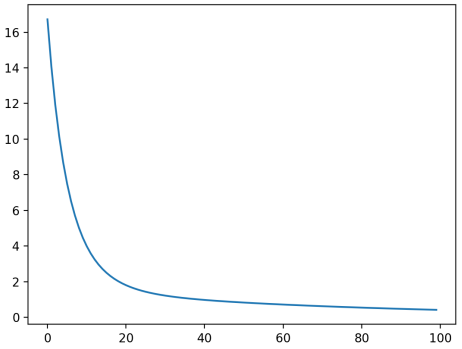
α = 0.001



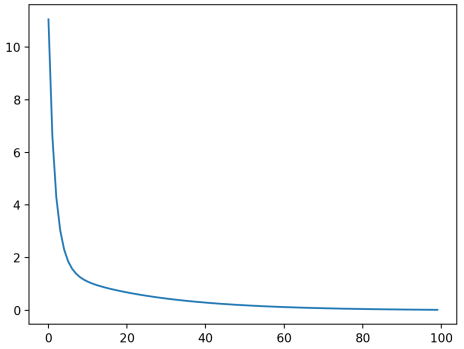
α = 0.1



α = 0.01

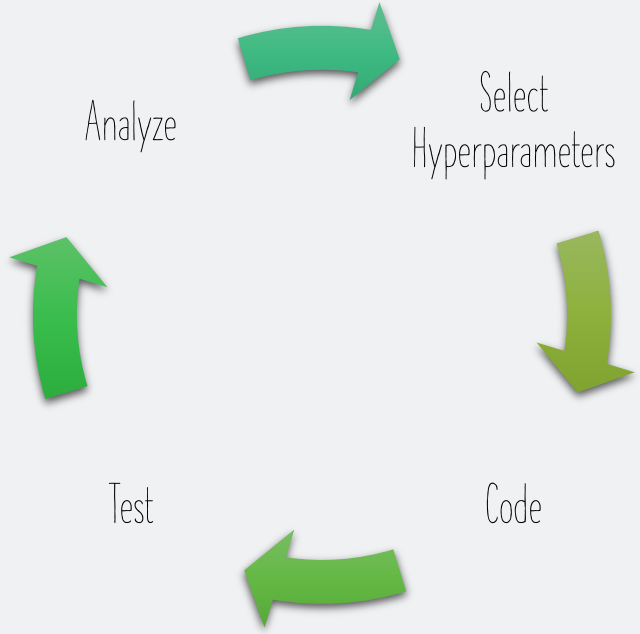


α = 0.3



Loss vs Epochs

FINE TUNNING



TIPS

Learning rate (alfa)

One of the most important hyperparameters

Use different learning rates

Record

- Loss
- Time (epochs)
- Accuracy

Scaling

Run without scaling

Use different scalings

- Normalization (different types)
- Standardization

BATCH SIZE

BATCH SIZE

Batch size usually limited by RAM + GPU

Number of instances that are being passed to the model at each iteration

Forward and Backward propagation

After each batch, the model parameters are updated

Recall that the number of batches are equal to the number of iterations

1 Epoch = all batches propagated

Finding the right size of the batch takes some time

BATCH SIZE...

Small batch sizes

- A less accurate gradient descent, as only part of the data is being used
- Finding of minima will take more time
- Slow training times
- Converge faster

Large batch sizes

- More accurate gradient descent calculation
- Faster progress in training
- Converge slowly
- Very large batch sizes have been found to have difficulty generalizing

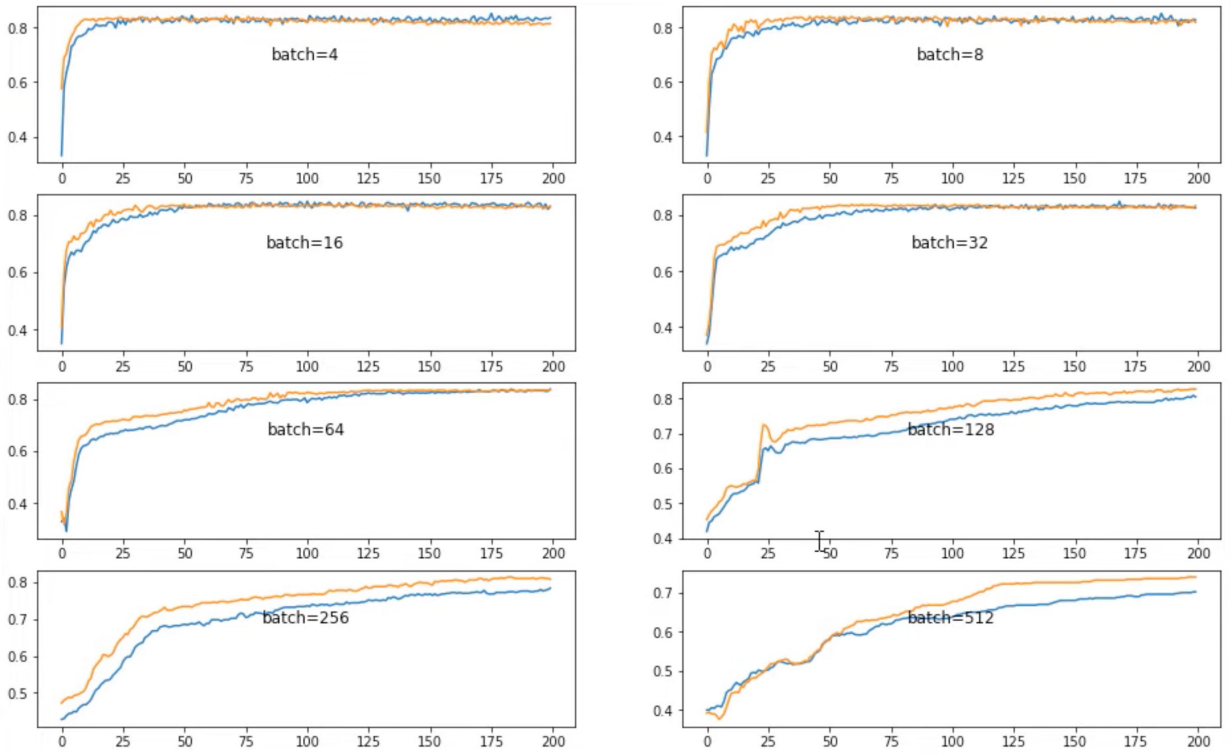
BATCH SIZE...

Batch size comparison using random data points

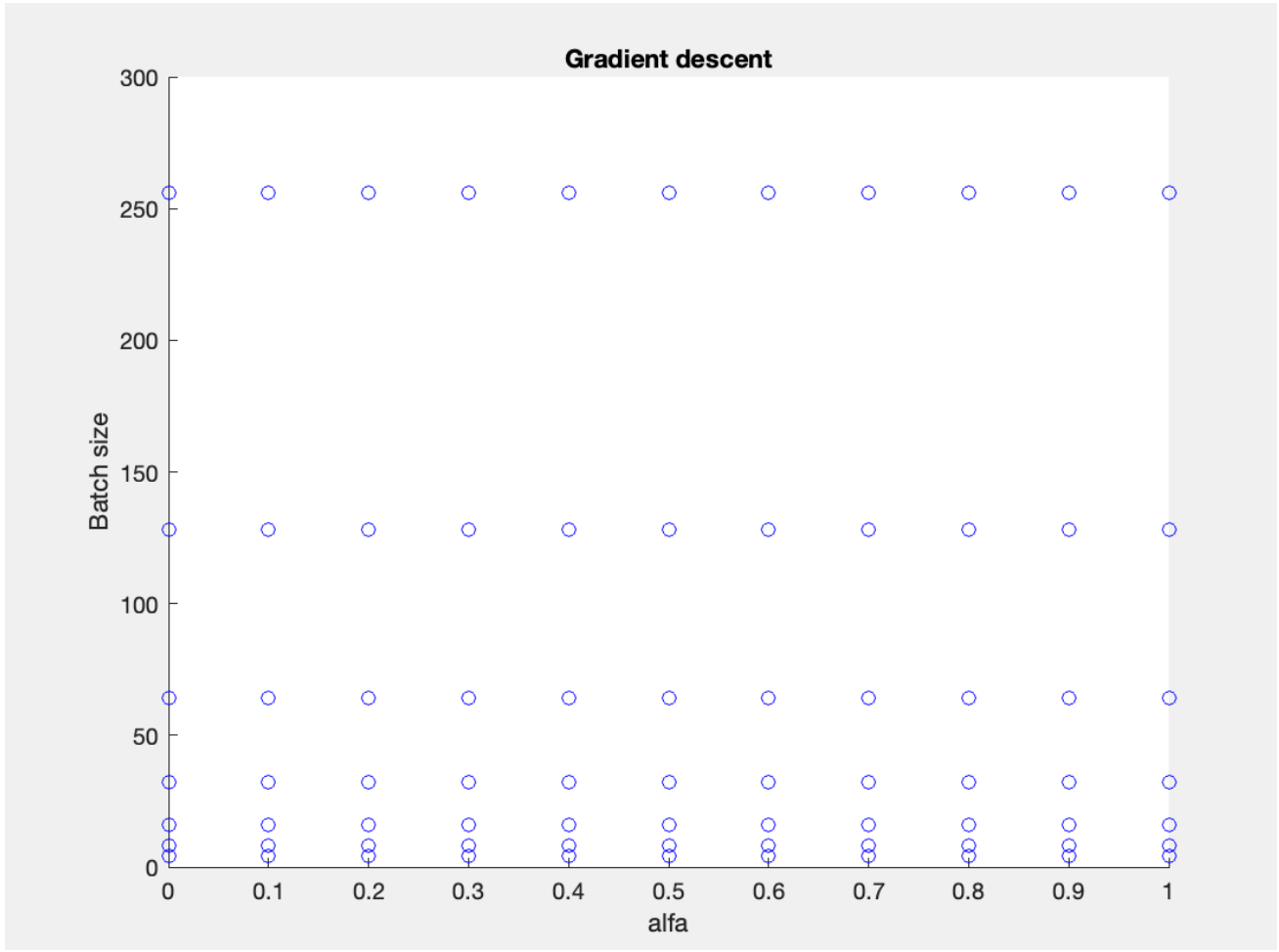
Smaller batches took a lot of time for training, but converged faster

Larger batches took very little time for training, but some failed to converge

Best sizes are usually 32 and 64



HYPERPARAMETER TUNING



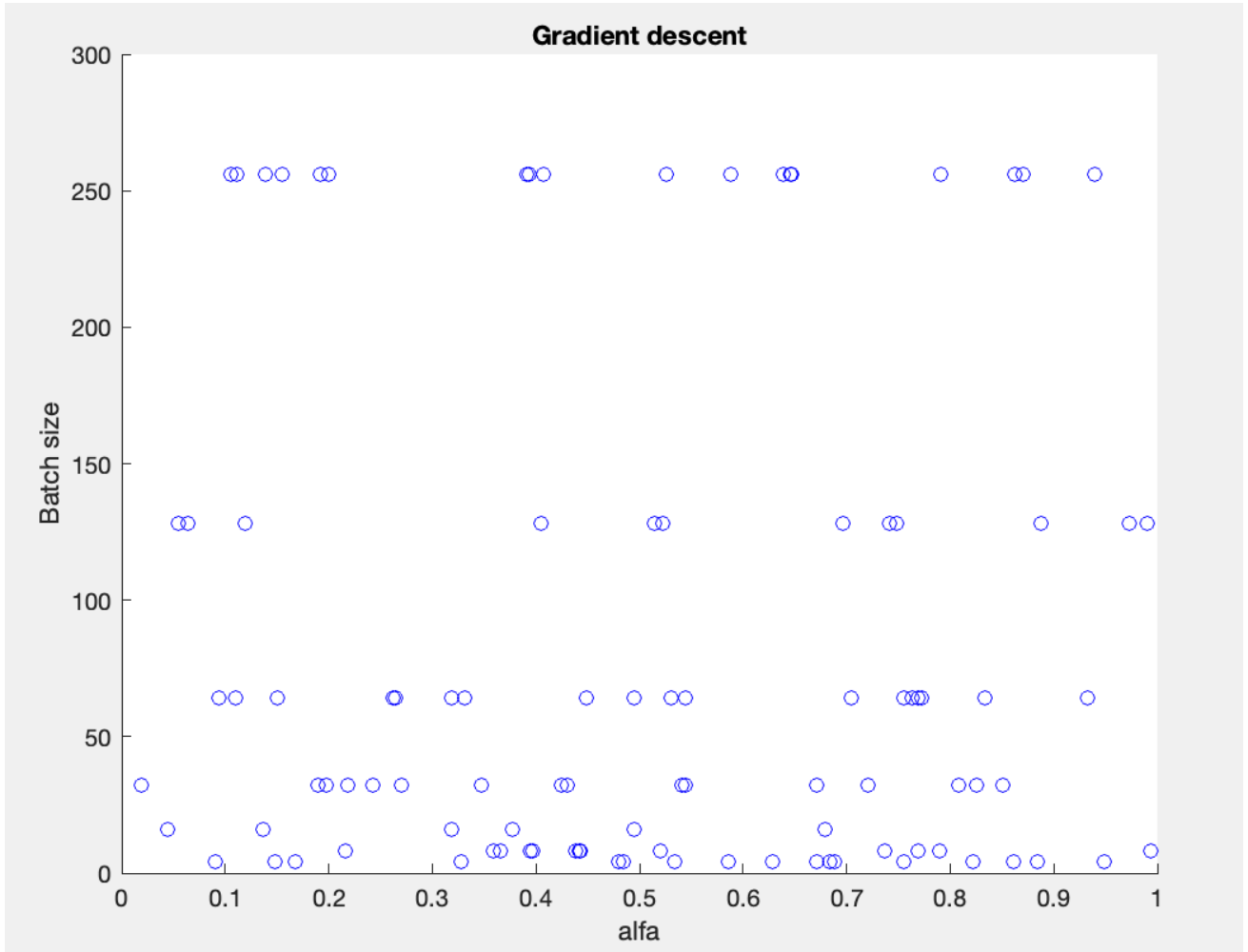
GRID SEARCH

Hyperparameter values are arranged in a grid

Values are explored systematically

Best combination is selected

Works well, but does not explore more combinations given that values are fixed



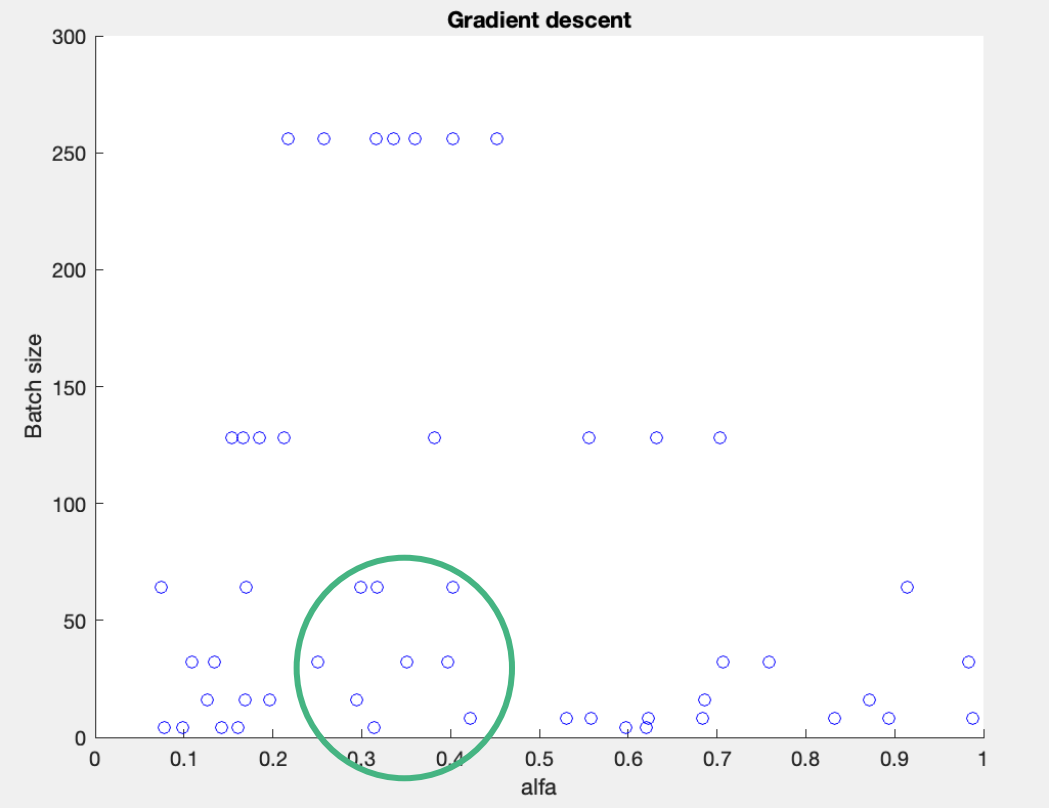
RAMDOM SEARCH

Try out all combinations of hyperparameters

In this case alfa values are more, so search in alfa is better

Search in Batch Size is the same as in Grid Search

COARSE TO FINE

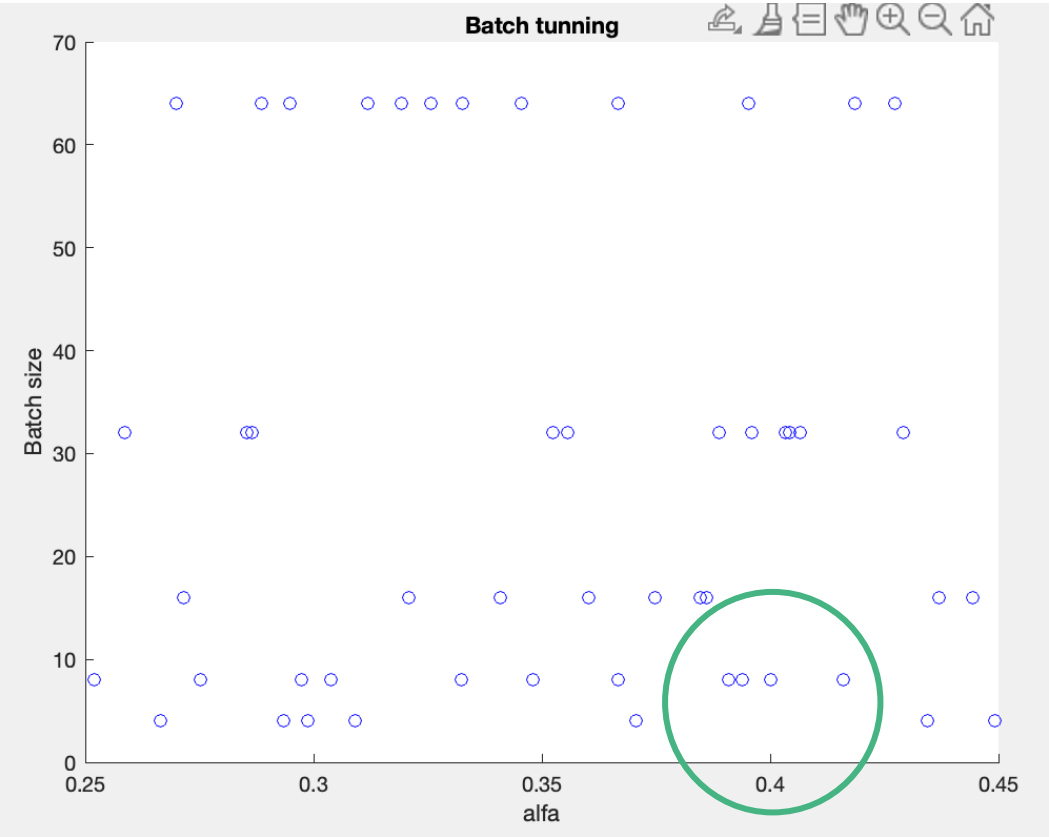


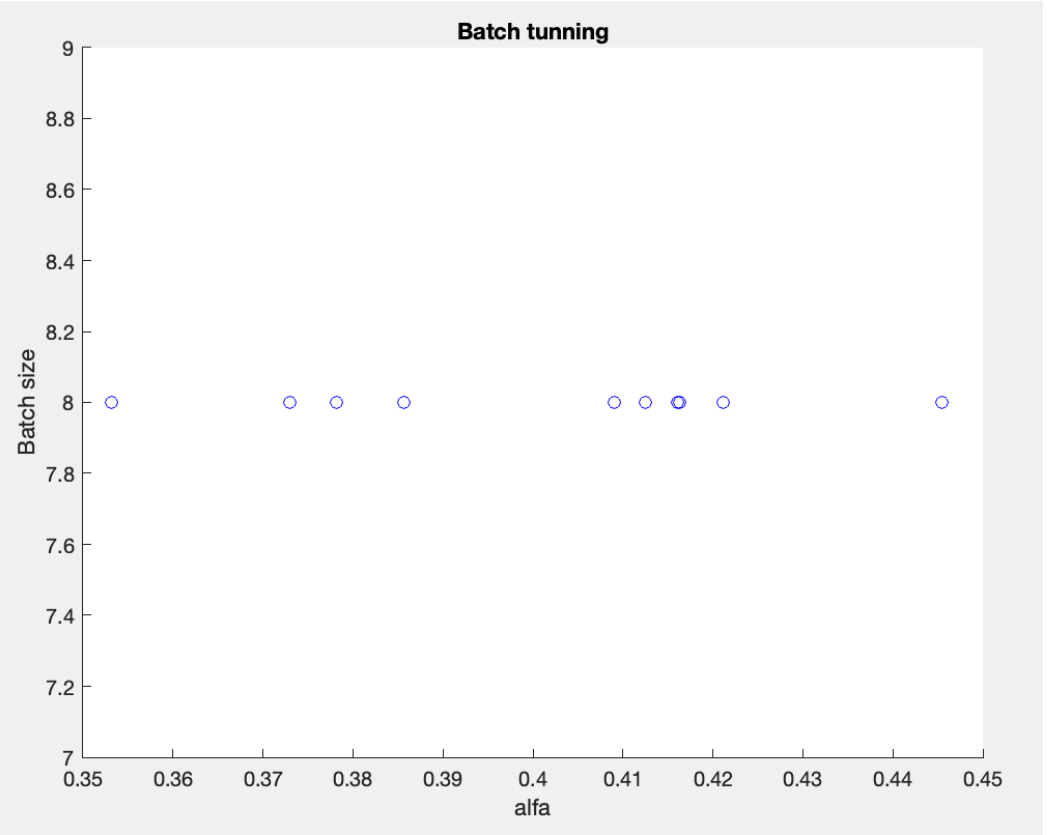
Select some initial values

Verify which values are the best

Fine tune in that area

COARSE TO FINE



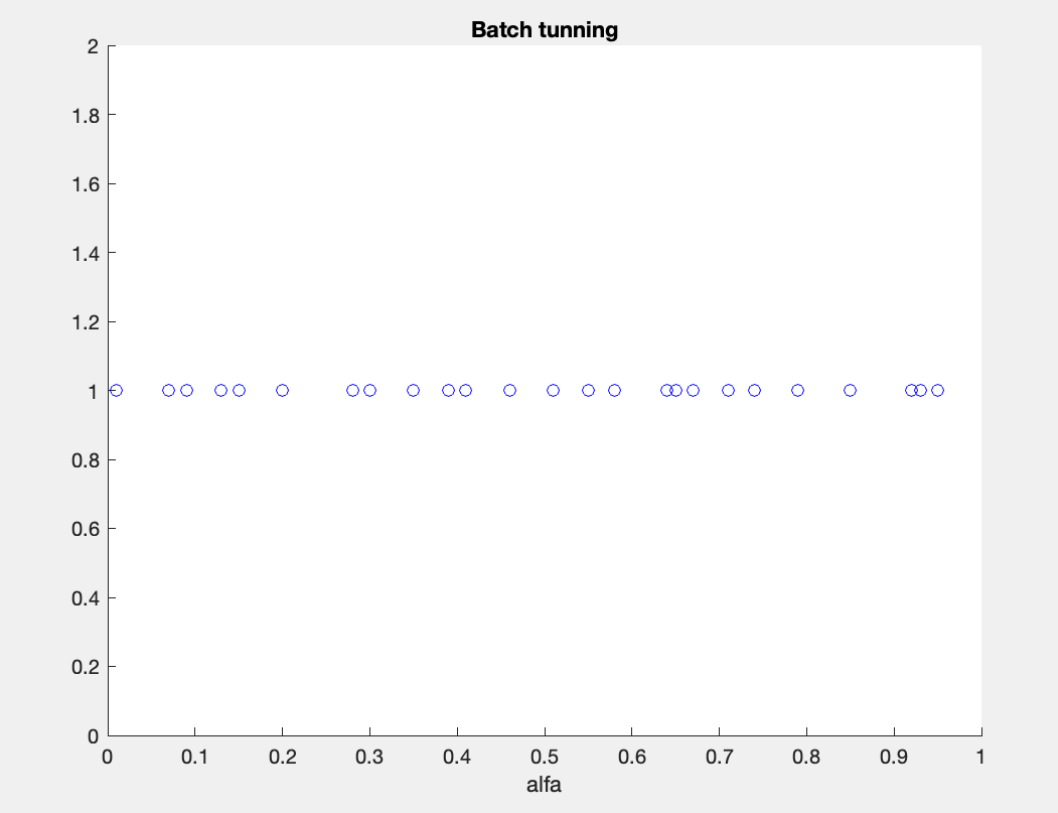


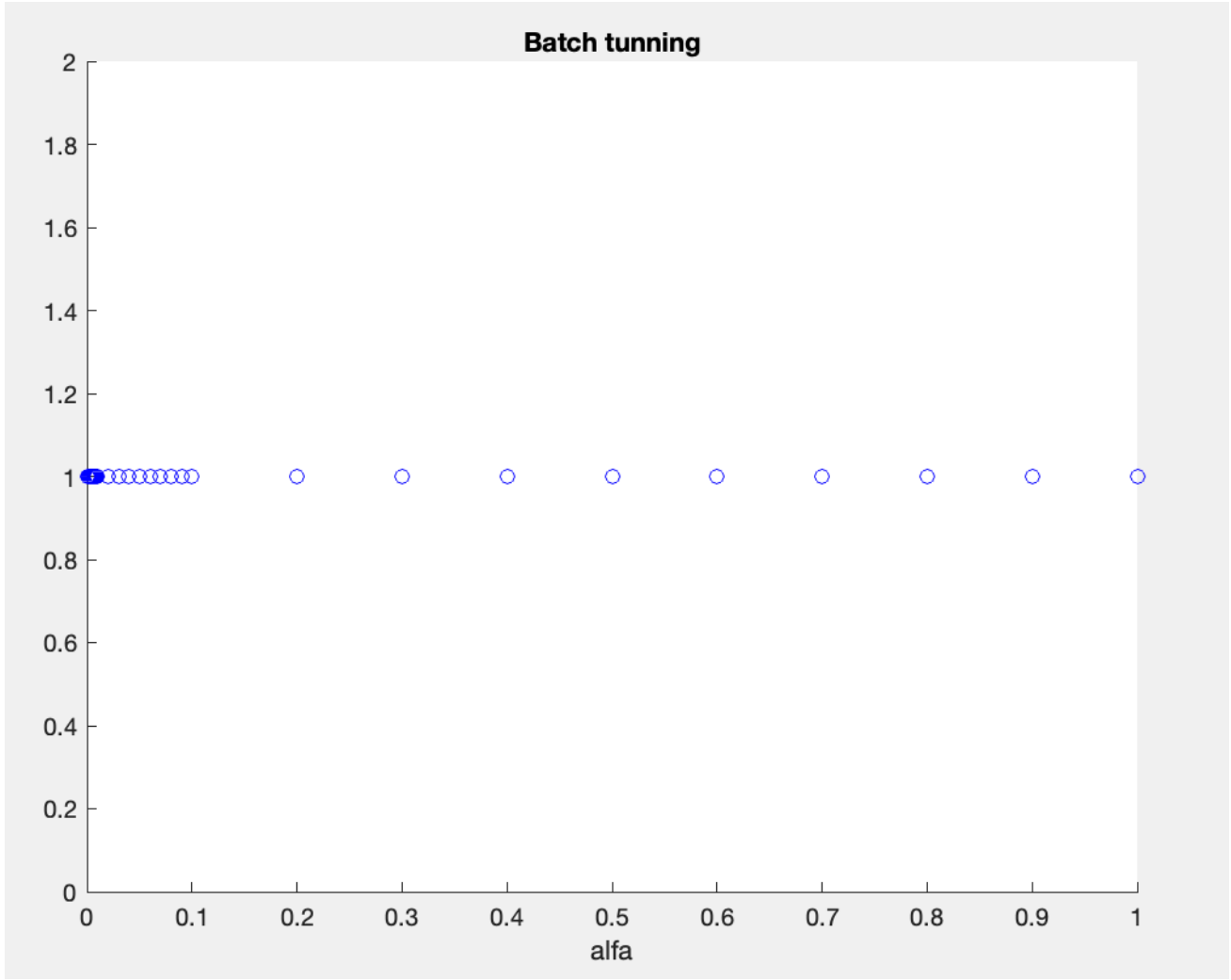
APPROPRIATE SCALE

Some hyperparameters need to be explored in more detail in some parts of their range

For example: learning rate

- Usually, best values below 0.5
- Values near 0 are best than values near 1





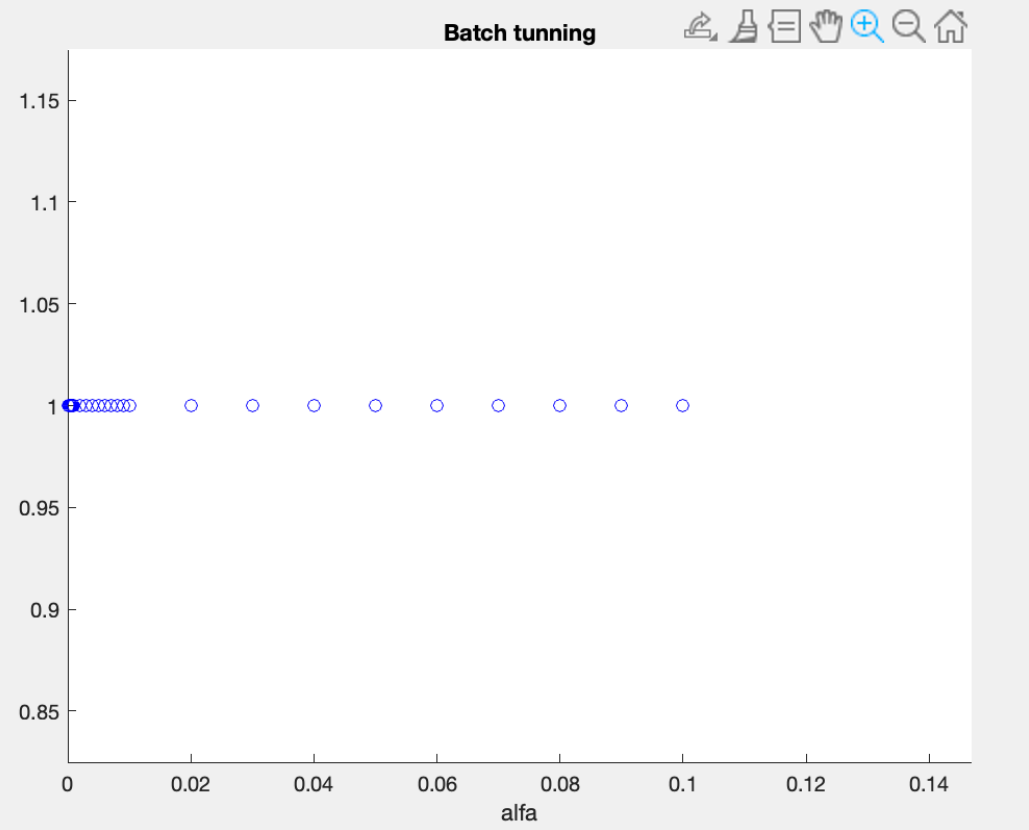
APPROPRIATE SCALE...

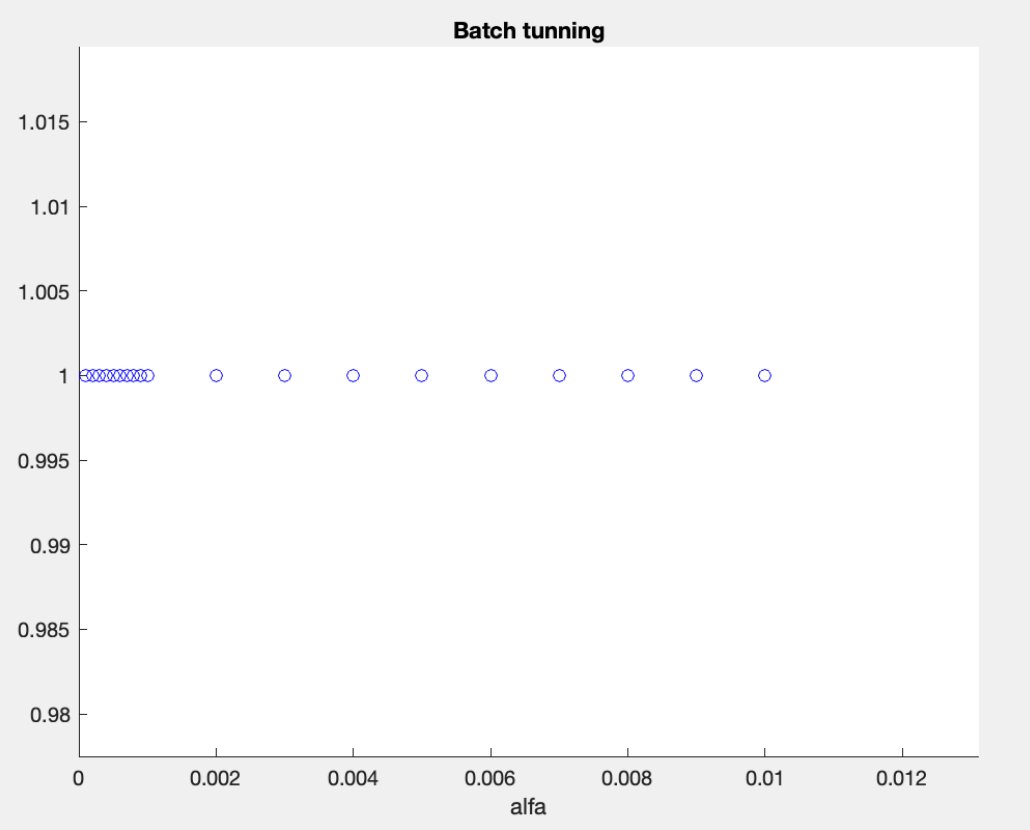
Best to search in a log scale

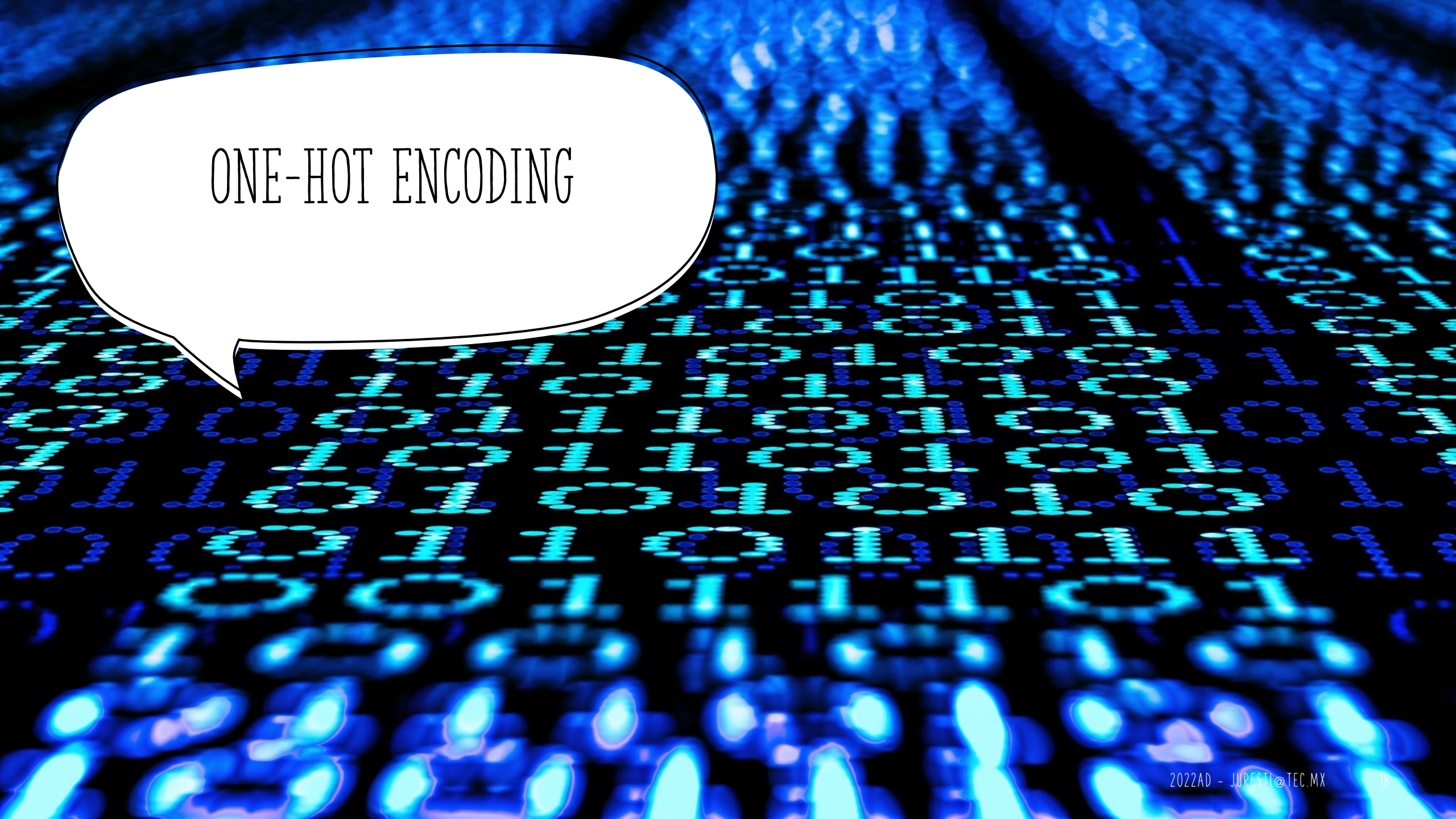
For example, learning rate:

- 0 to 0.0001
- 0.0001 to 0.001
- 0.001 to 0.01
- 0.01 to 0.1
- 0.01 to 1

APPROPRIATE SCALE...






ONE-HOT ENCODING

ONE-HOT ENCODE

Categorical or nominal data

Data that does not have a numerical value

It usually does not have a natural ordering of values

Not all ML methods can work with categorical data

One-hot encode

Prepares categorical (nominal) data for processing

Converts categorical data to numerical data

Easily reversible

INTEGER ENCODING

For categorical variables that have a natural order

Example:

Size = [small, medium, large]

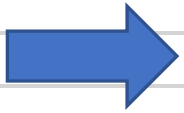
Each nominal value is assigned a numerical value (usually an integer)

Any numerical value that follows an order is allowed

Original data	Integer encoding
Small	1
Medium	2
Large	3

ONE-HOT ENCODE

Color		Red	Green	Blue	Black
Red		1	0	0	0
Green		0	1	0	0
Blue		0	0	1	0
Black		0	0	0	1
Red		1	0	0	0
Black		0	0	0	1
Green		0	1	0	0



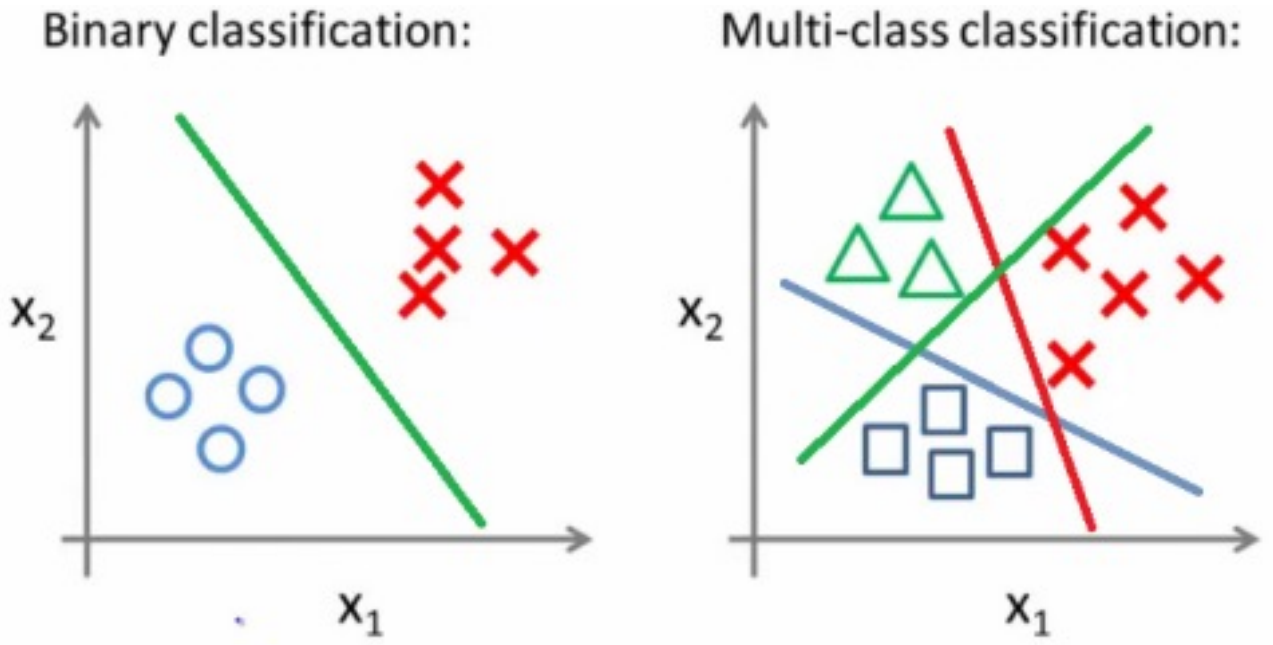
For categorical variables where there is no natural order

A binary variable is added for each unique value of the nominal variable

Example:

Color = [Red, Green, Blue, Black]


EXTENSIONS OF LOGISTIC REGRESSION



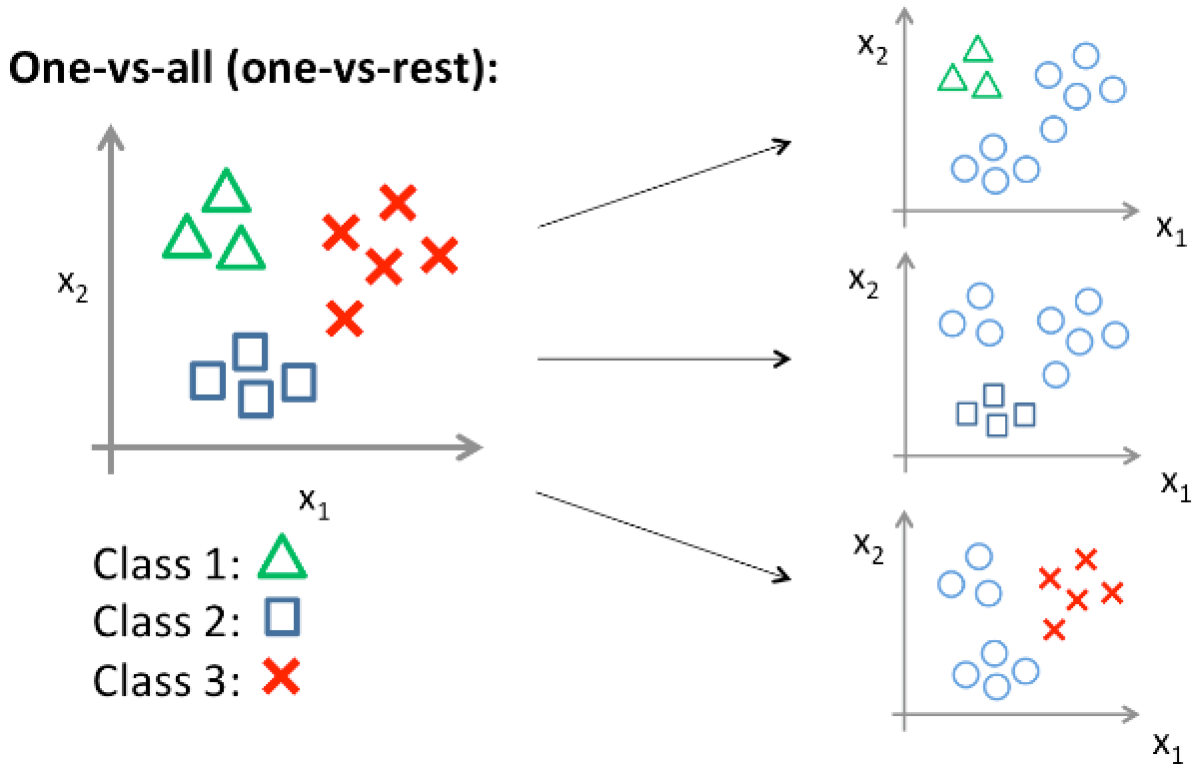
https://utkuufuk.com/2018/06/03/one-vs-all-classification/one-vs-all.png

ONE VS. ALL (ONE-VS-REST) STRATEGY

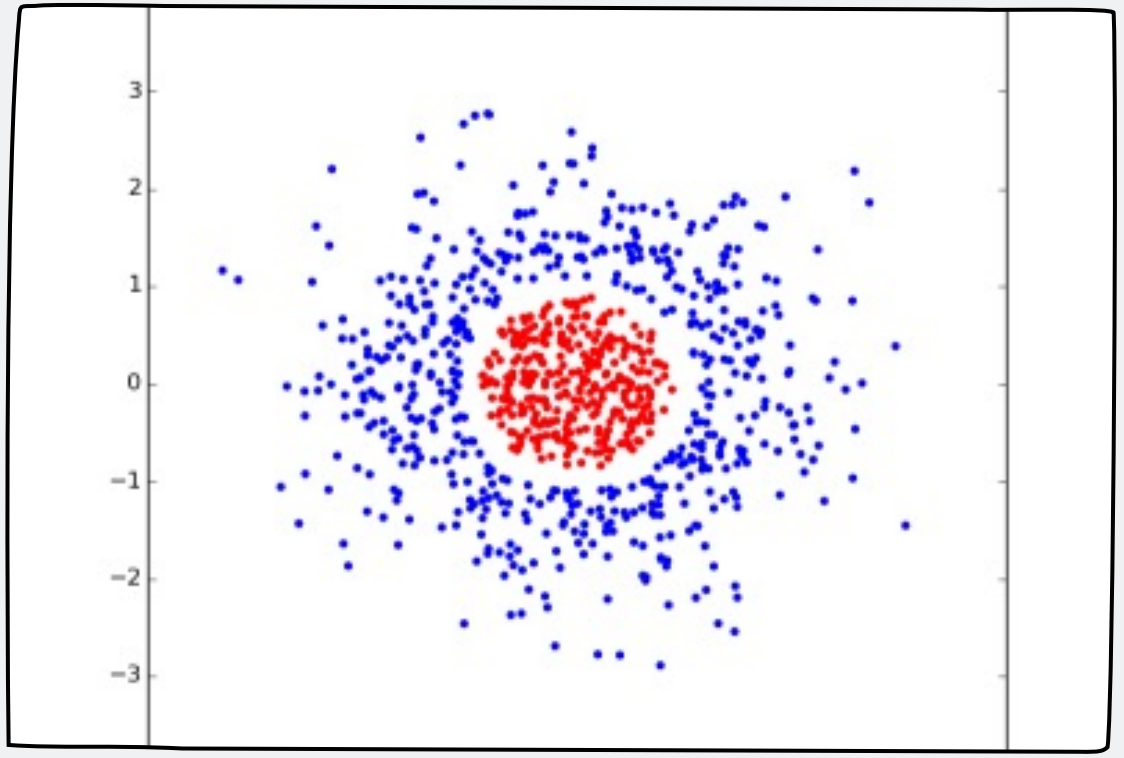
Binary classifier can only classify 2 classes

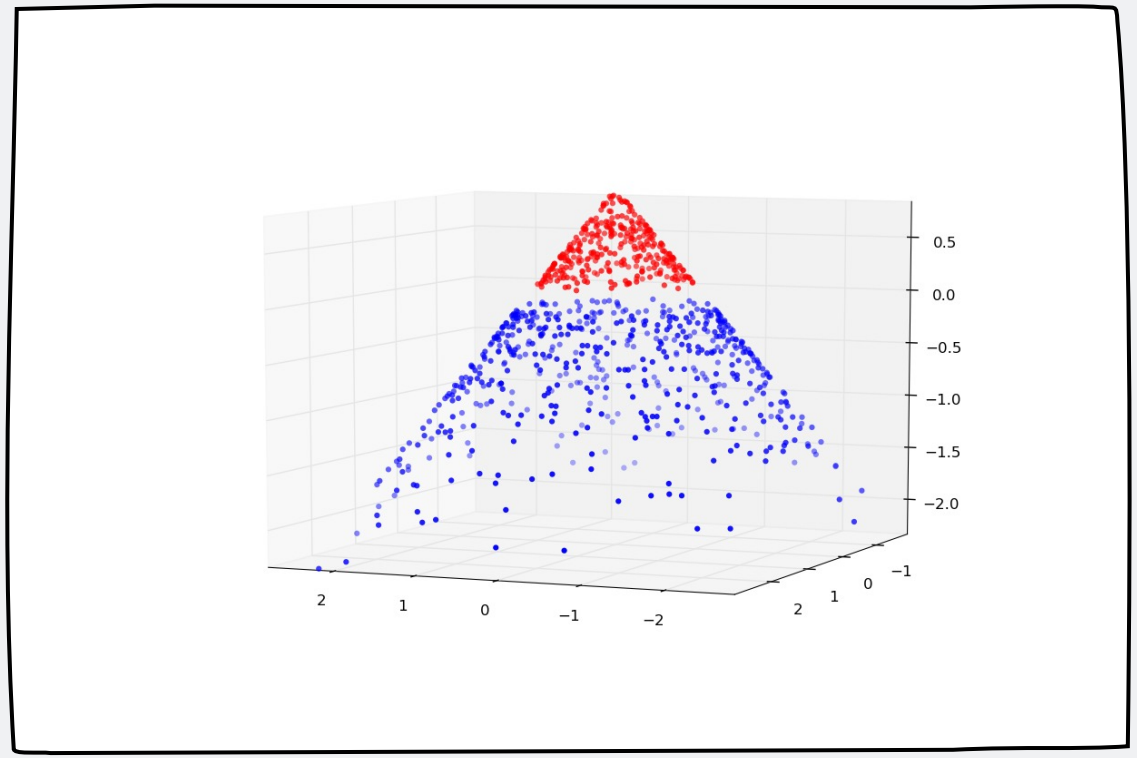
Multiclass classification is done by classifying N classes with N binary classifiers

ONE VS. ALL (ONE-VS-REST) STRATEGY...



https://www.cc.gatech.edu/classes/AY2016/cs4476_fall/results/proj4/html/jnanda3/svm.png





CHANGE OF DIMENSION