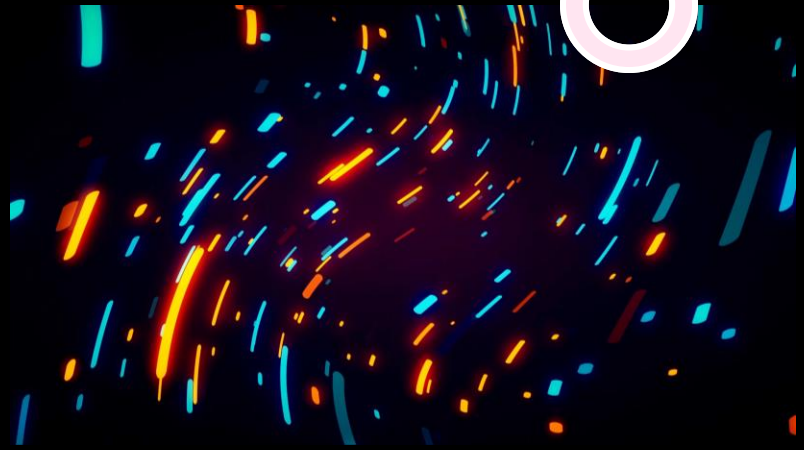
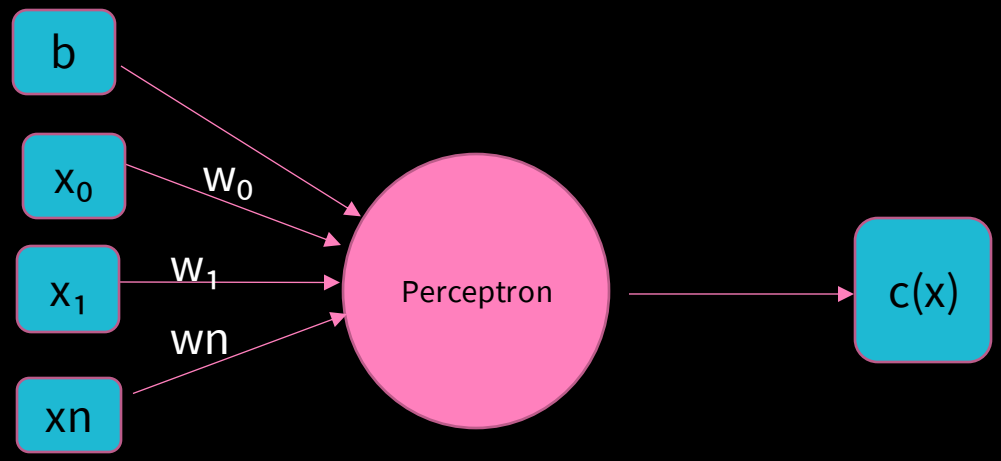
PERCEPTRON LEARNING ALGORITHM

- KHALED QURAINIS
- LIA MURPHY
- JULIO OJALVO



Perceptron Summary

- The idea of the Perceptron:
  - ➢ Inputs
  - ➢ Weights
  - ➢ Perceptron
  - ➢ Output



- $c(x) = b + x_0 * w_0 + x_1 * w_1 \ldots\ldots + xn * wn$

Weight Updates

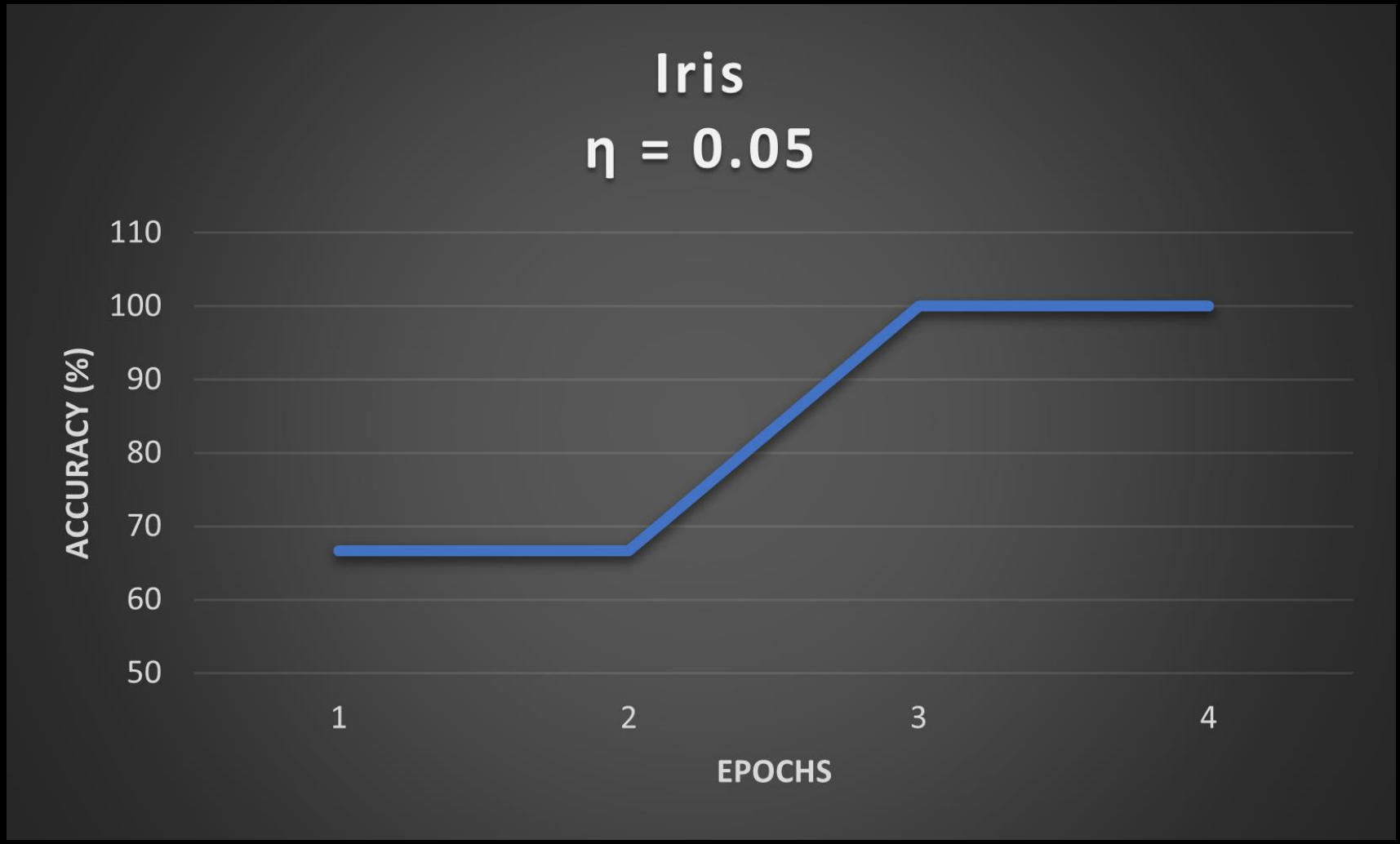
- Calculate the hypothesis h(x) of the example
  - Pos or Neg?
- If h(x) is the correct class, weights don’t change
- If h(x) is greater than the class, reduce weights
- If h(x) is less than the class, increase weights

- Weights are scaled by a learning rate (η)

Databases

- Iris dataset (150 examples)
  - 50 pos, 100 neg
- Balloons dataset (20 examples)
  - 8 pos, 12 neg
- Banknote dataset (1372 examples)
  - 610 pos, 762 neg

Iris Database



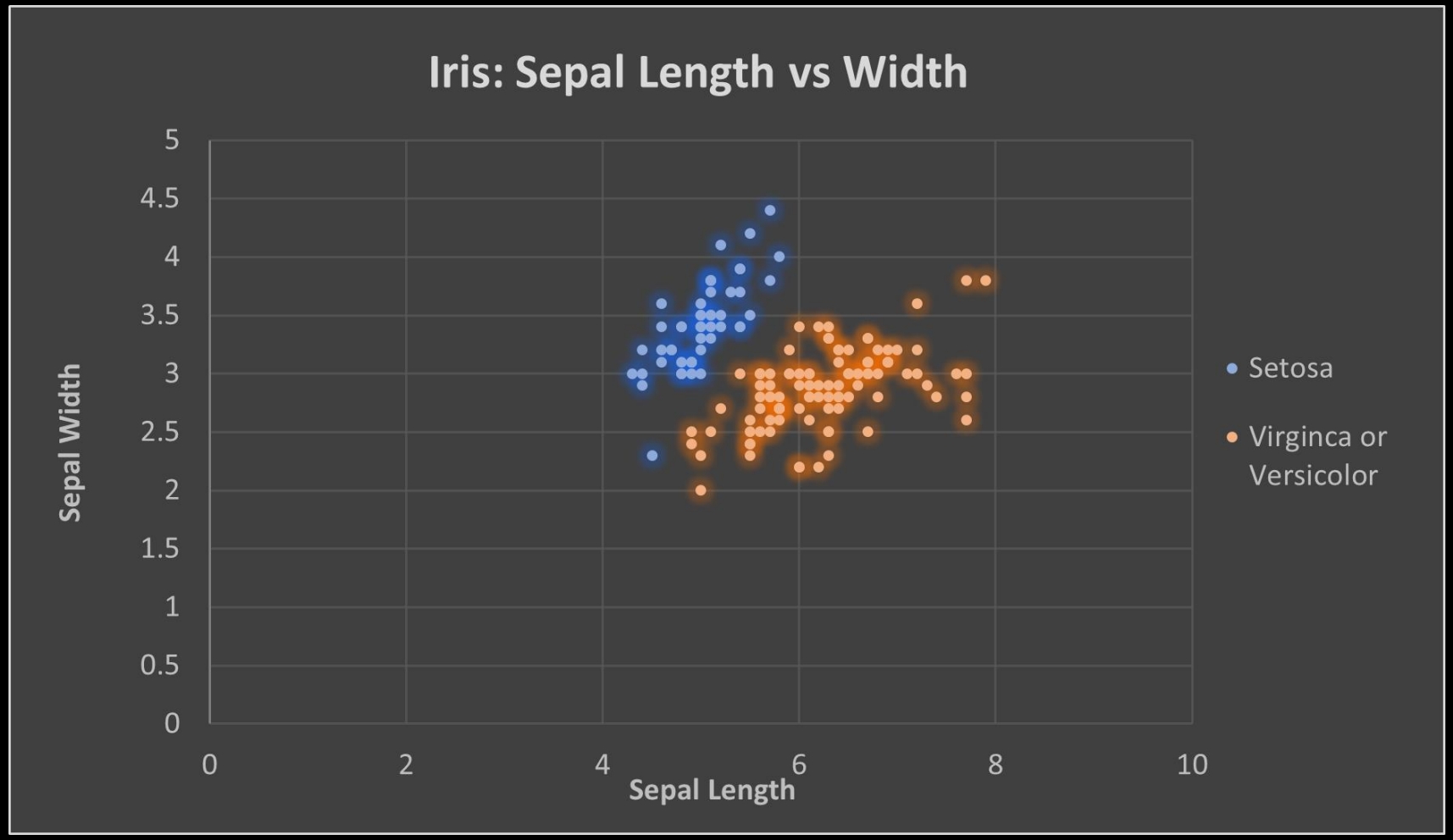
Iris Database

```
Epoch: 0
Weights[0] = -1.65
Weights[1] = 0
Weights[2] = -3.95
Weights[3] = -1.75

Epoch: 1
Weights[0] = 0.9
Weights[1] = 1.75
Weights[2] = -3.25
Weights[3] = -1.65

Epoch: 2
Weights[0] = 0.9
Weights[1] = 1.75
Weights[2] = -3.25
Weights[3] = -1.65

Epoch: 3
Weights[0] = 0.9
Weights[1] = 1.75
Weights[2] = -3.25
Weights[3] = -1.65

Epoch: 4
Weights[0] = 0.9
Weights[1] = 1.75
Weights[2] = -3.25
Weights[3] = -1.65
```

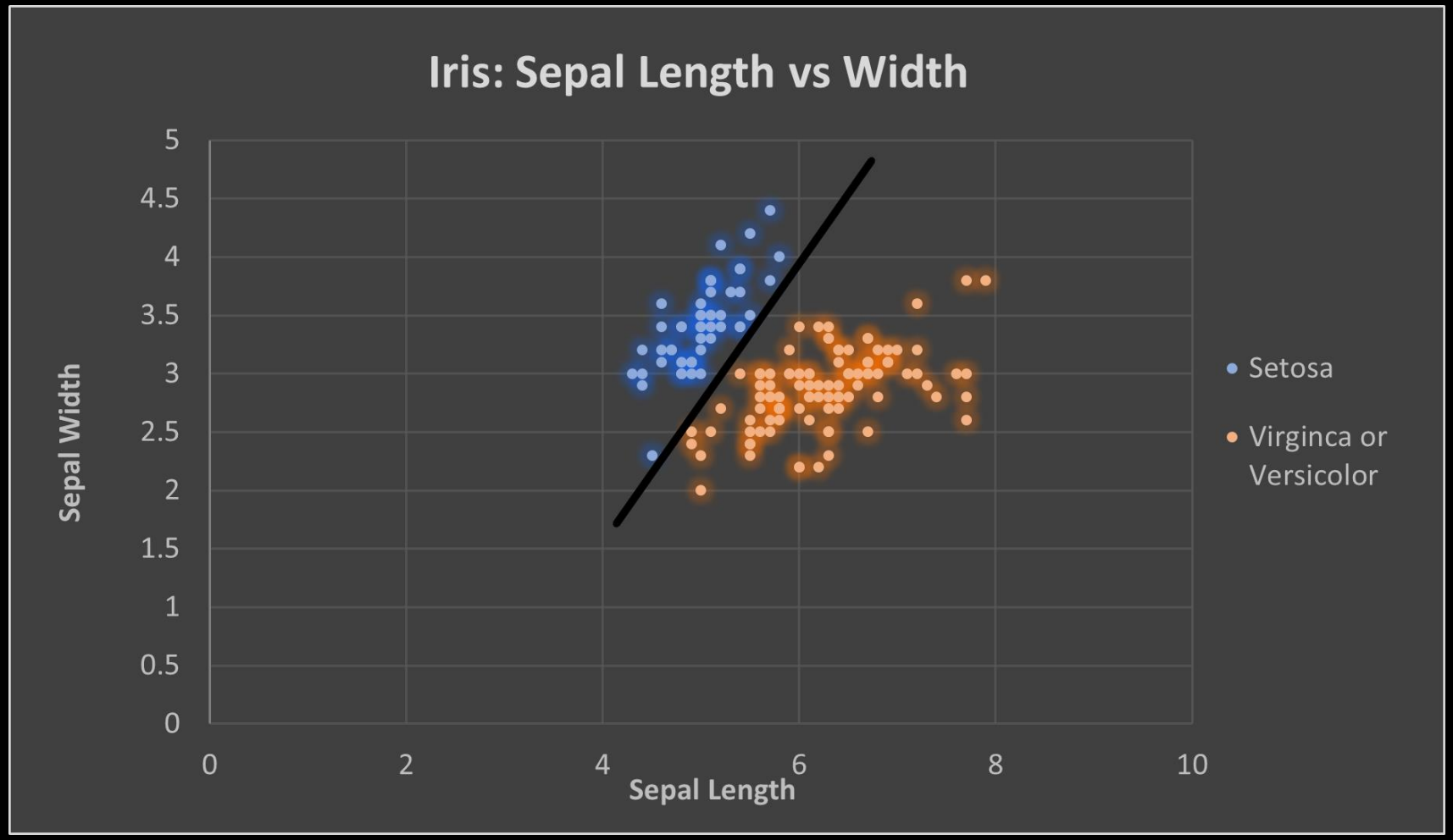
Final classifier:

$c(x) = 0.9x_0 + 1.75x_1 - 3.25x_2 - 1.65x_3$

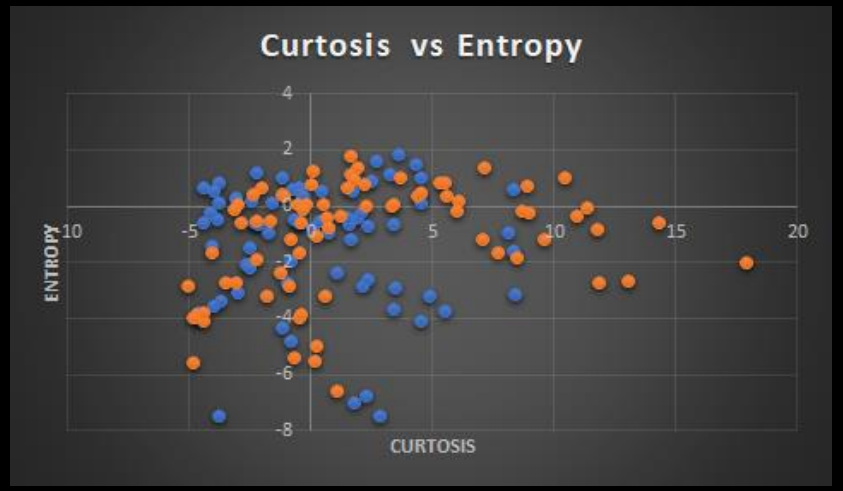
Linearly Separable Data

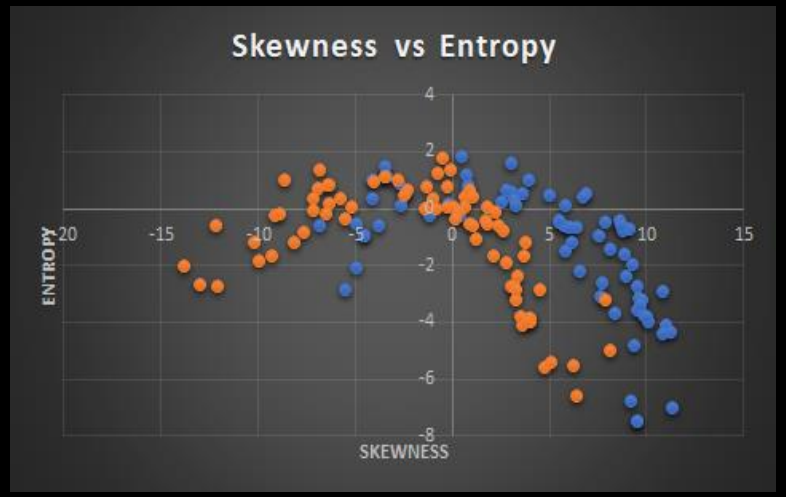


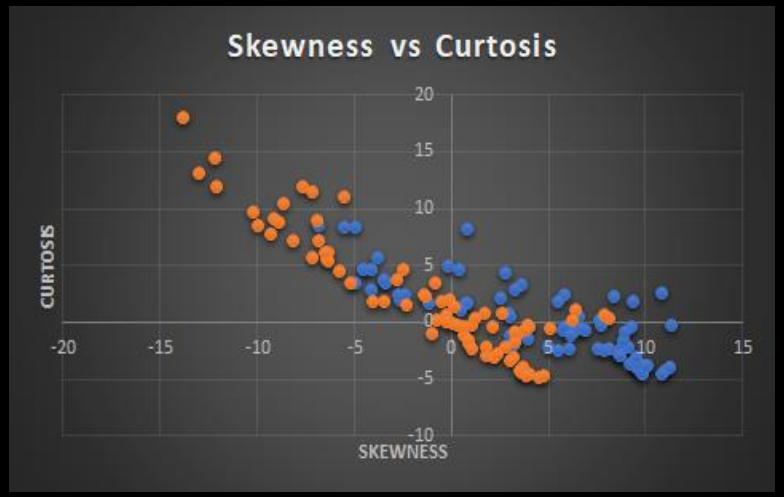
Linearly Separable Data

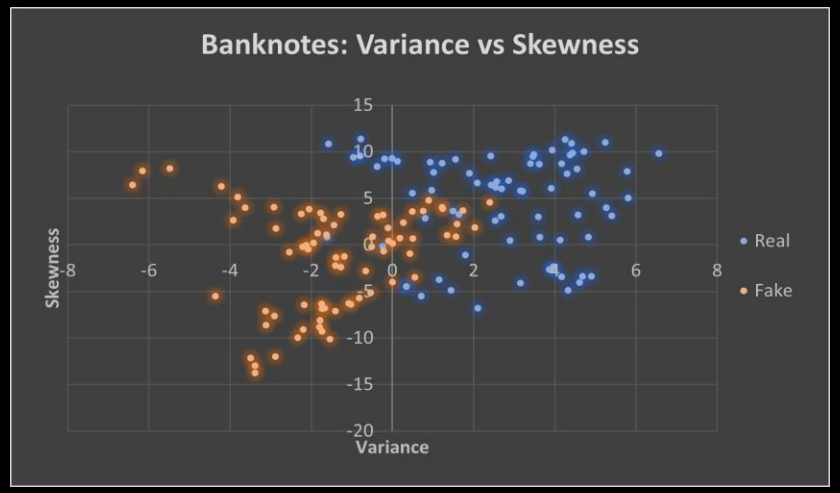


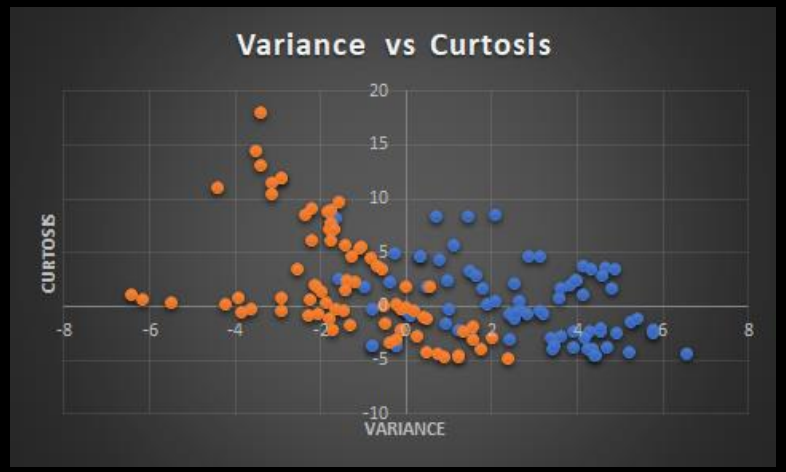
Non-Linearly Separable Data

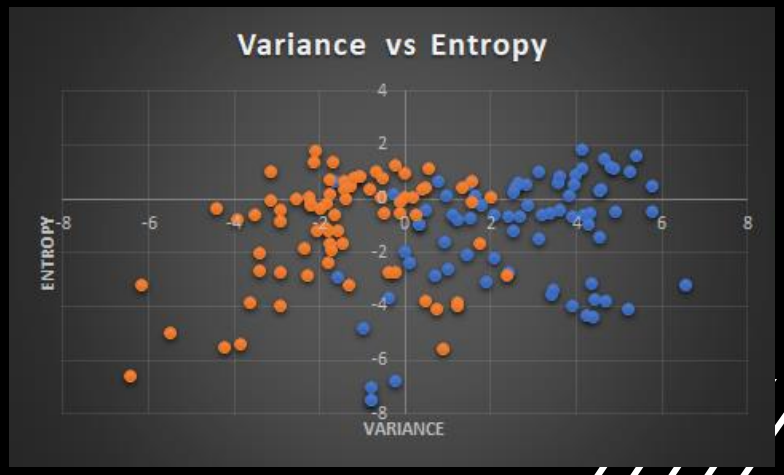






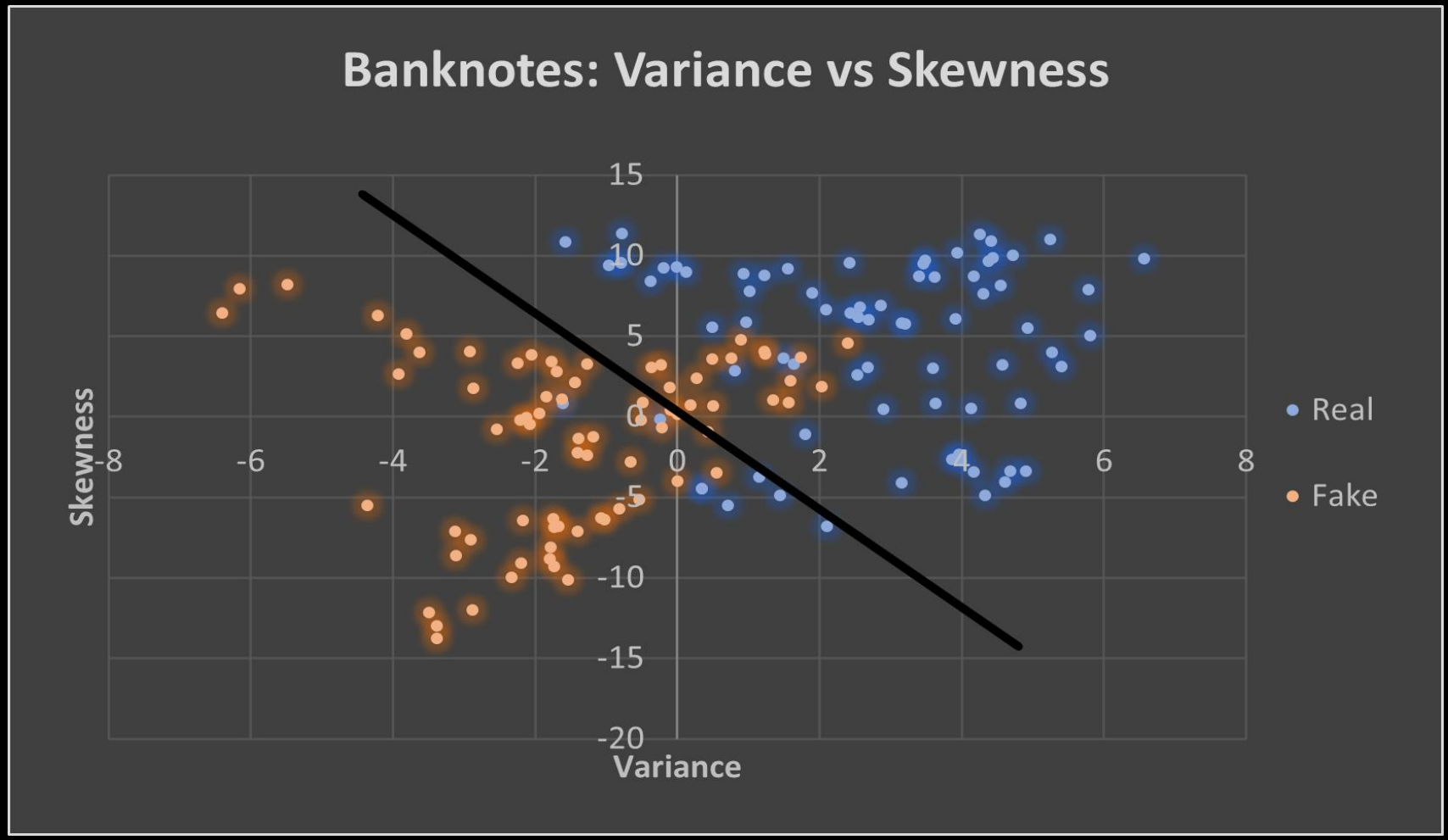




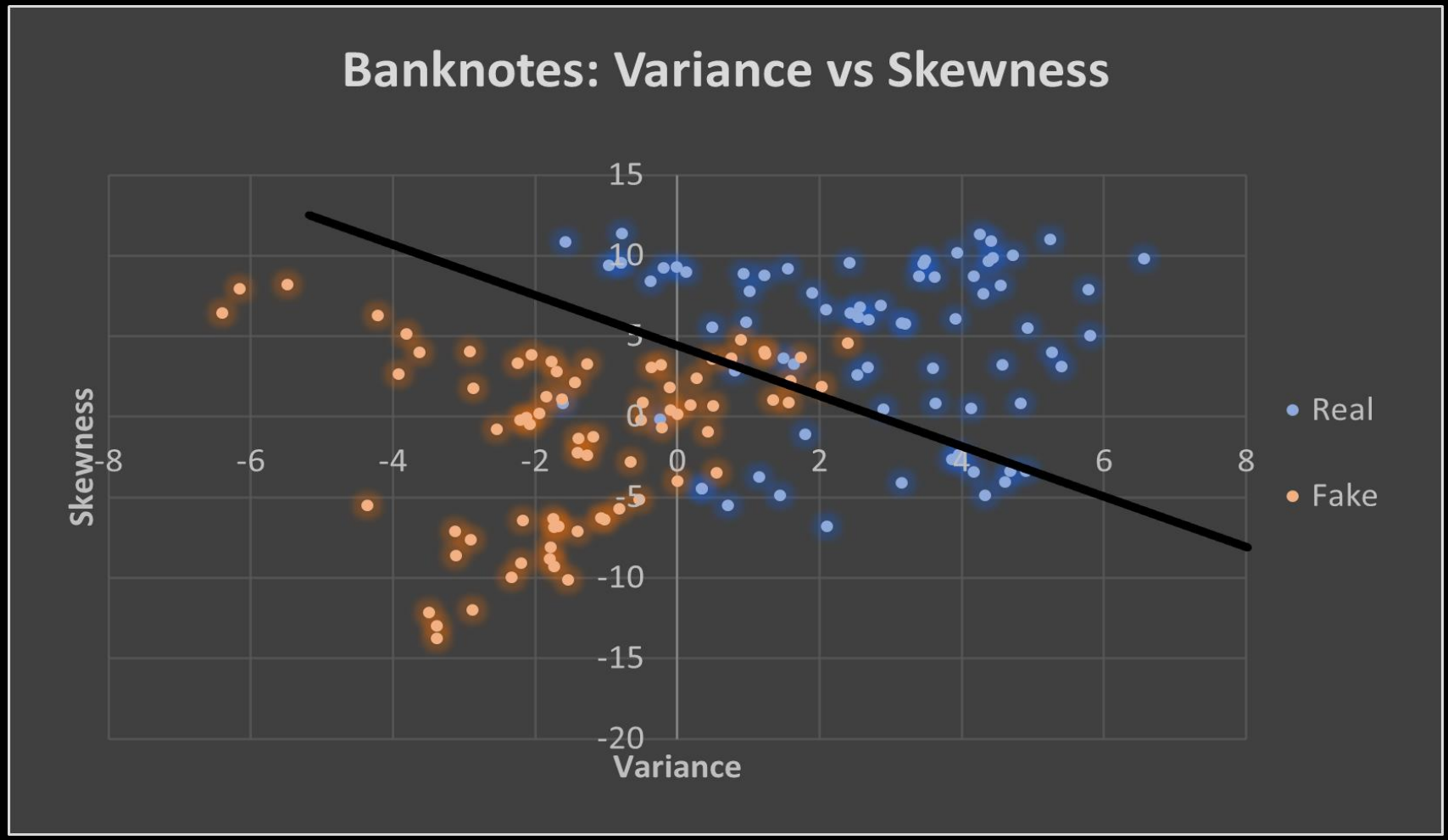


*number of examples = 150

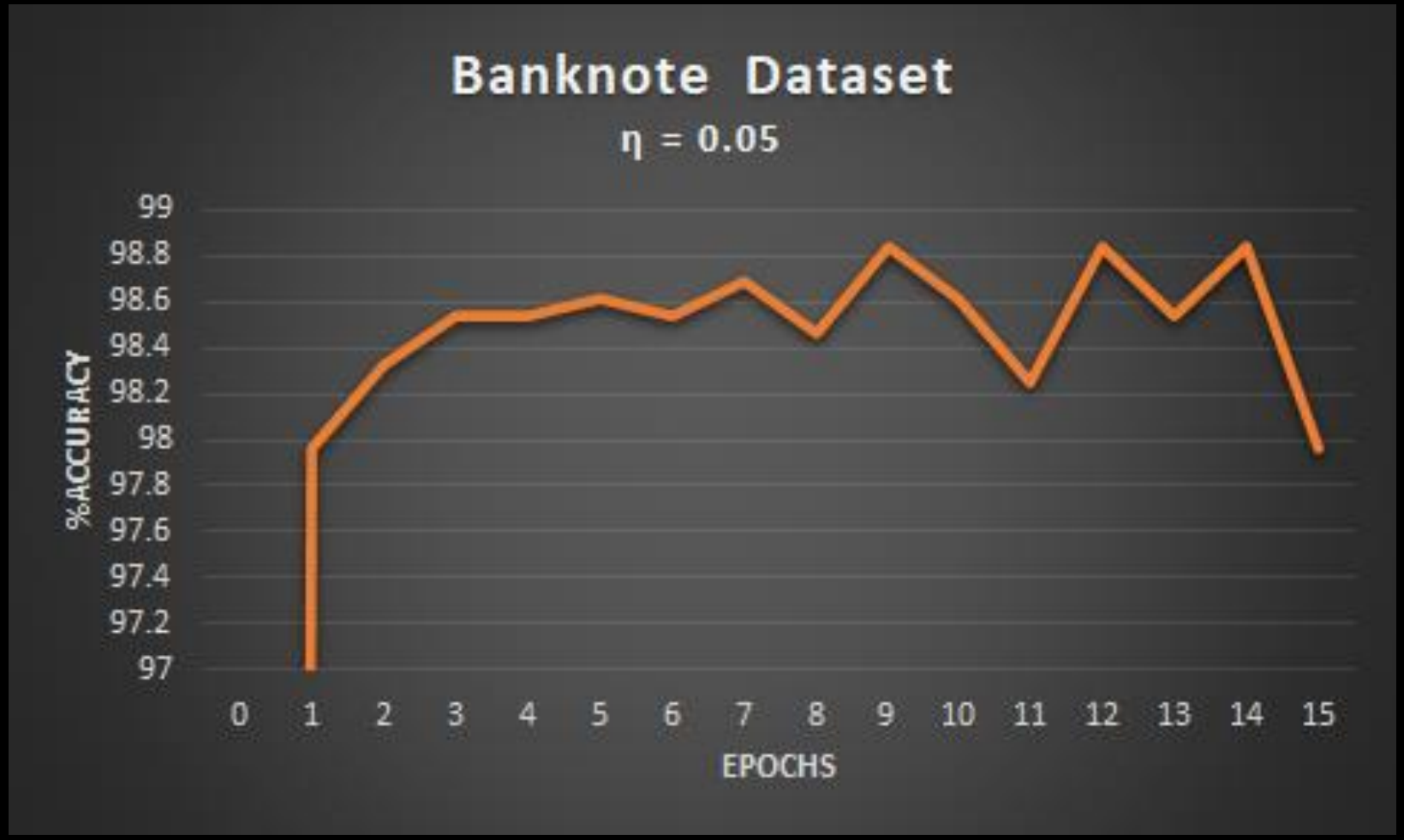
Non-Linearly Separable Data



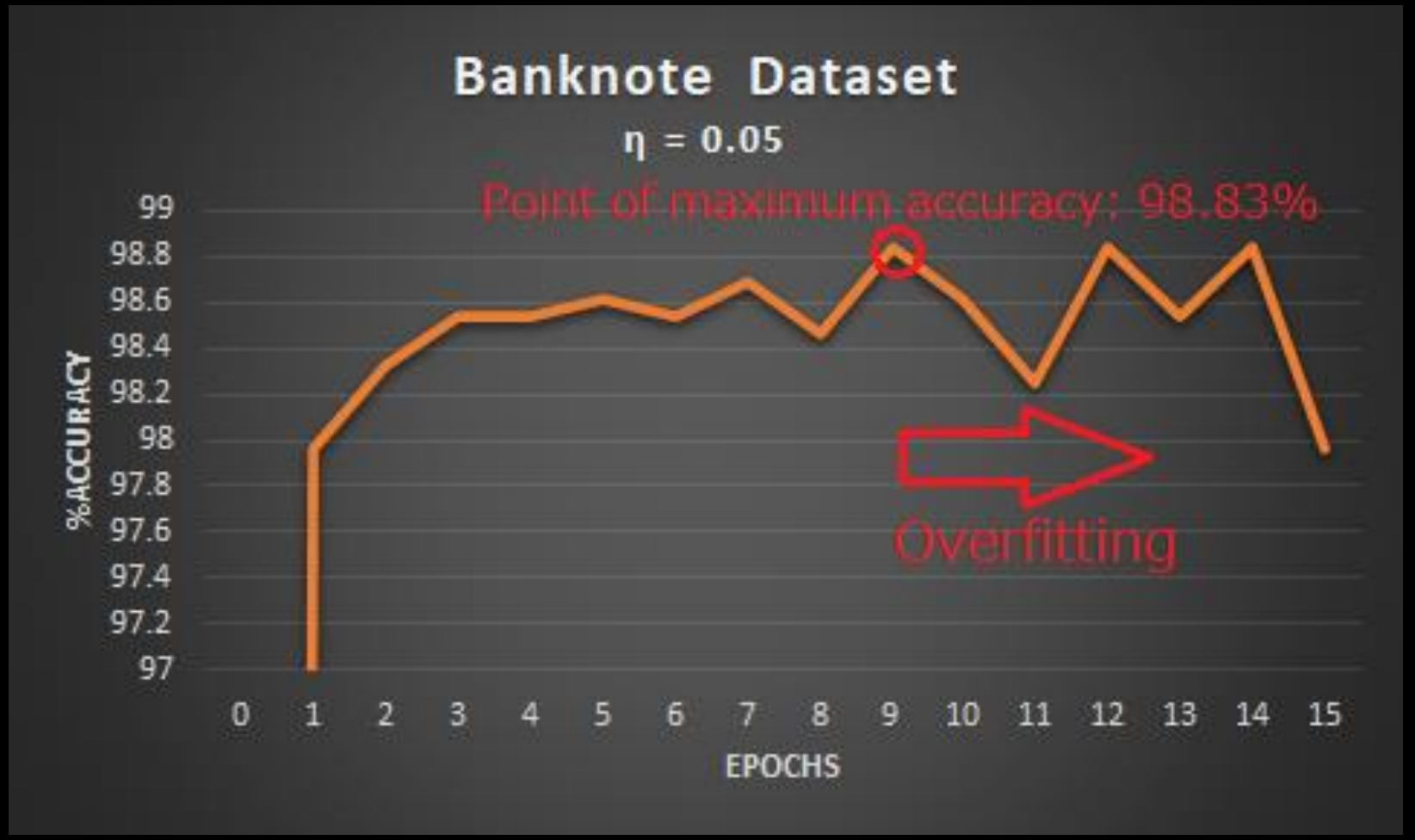
Non-Linearly Separable Data



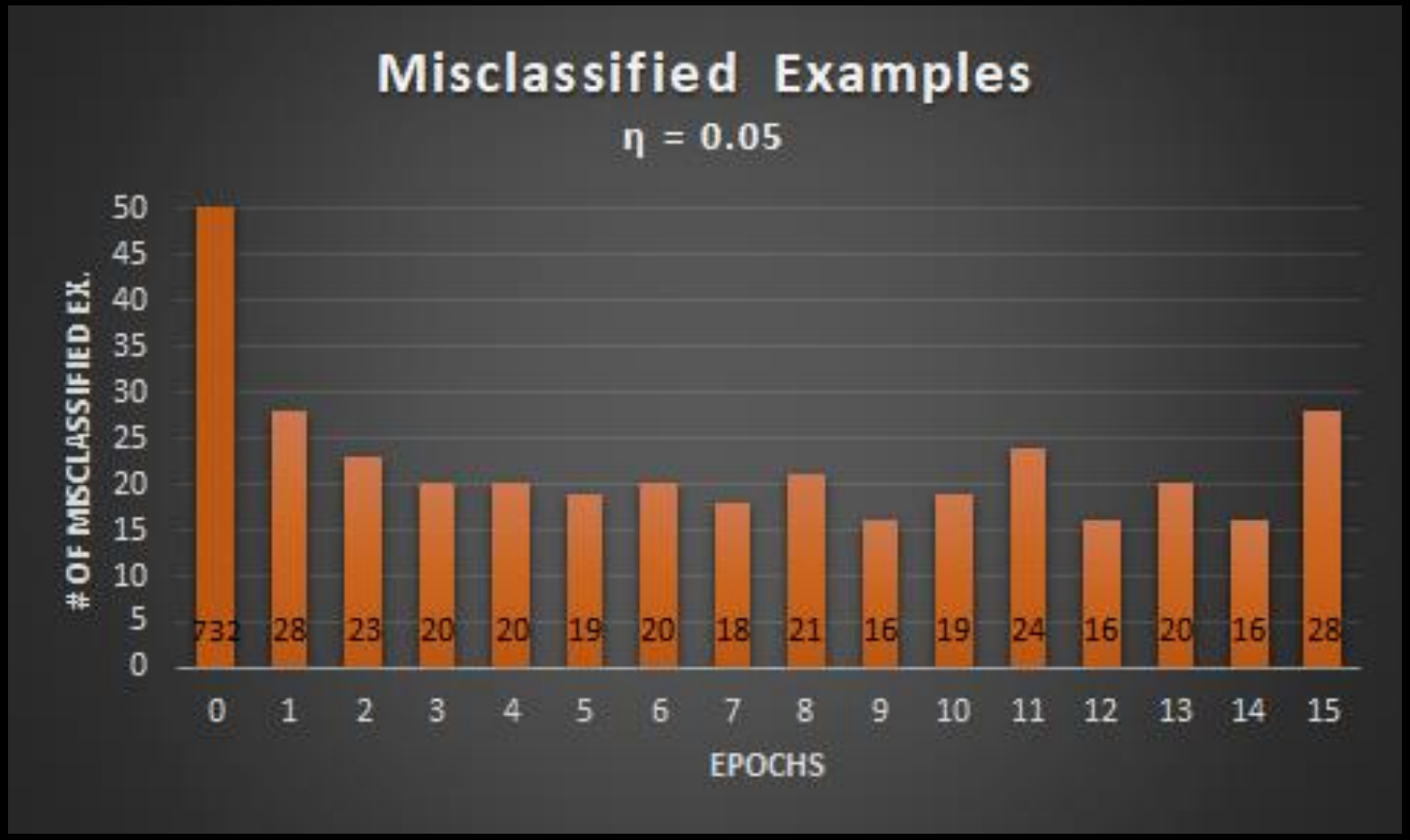
Banknote dataset



Banknote dataset



Banknote dataset

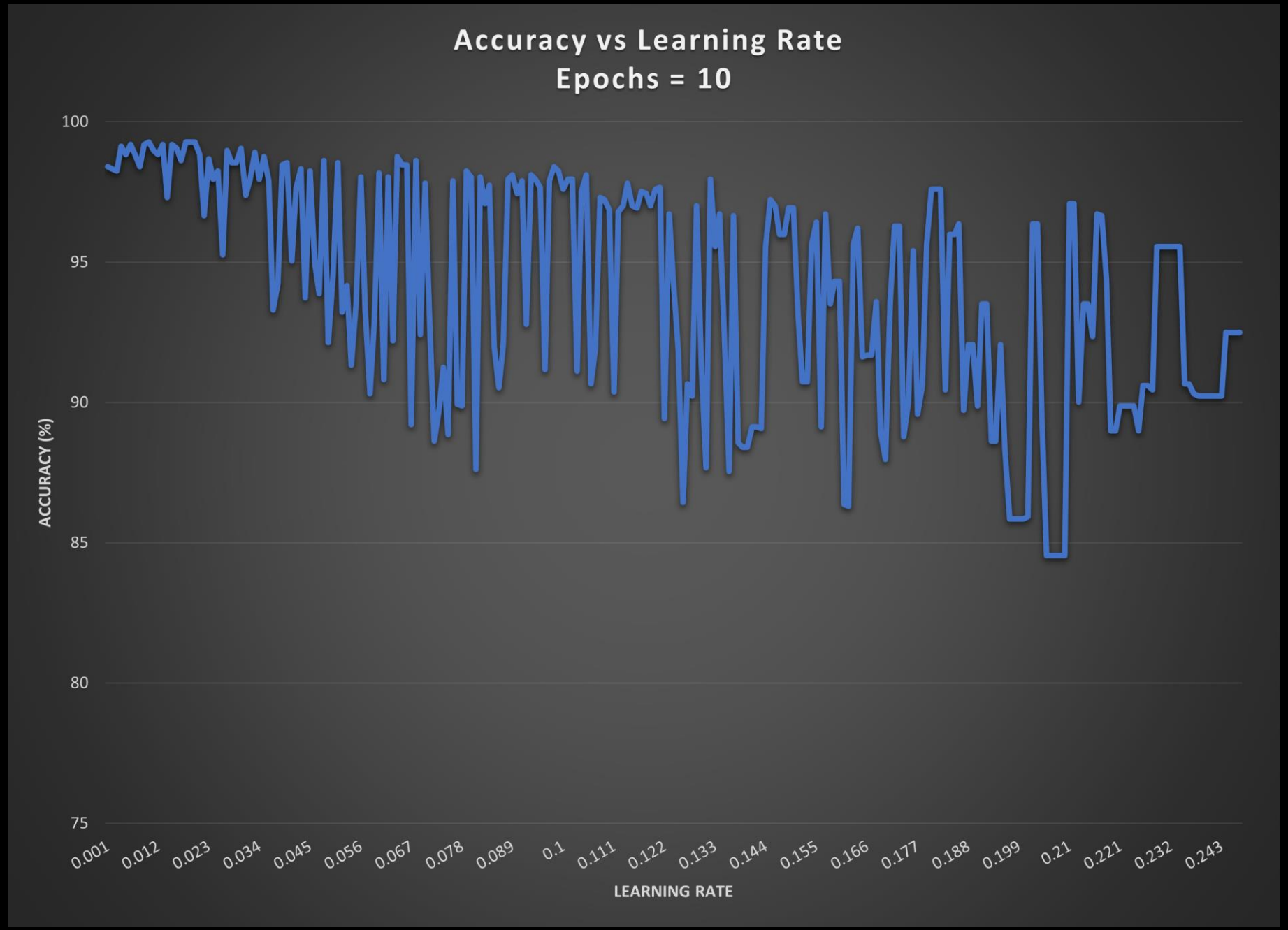


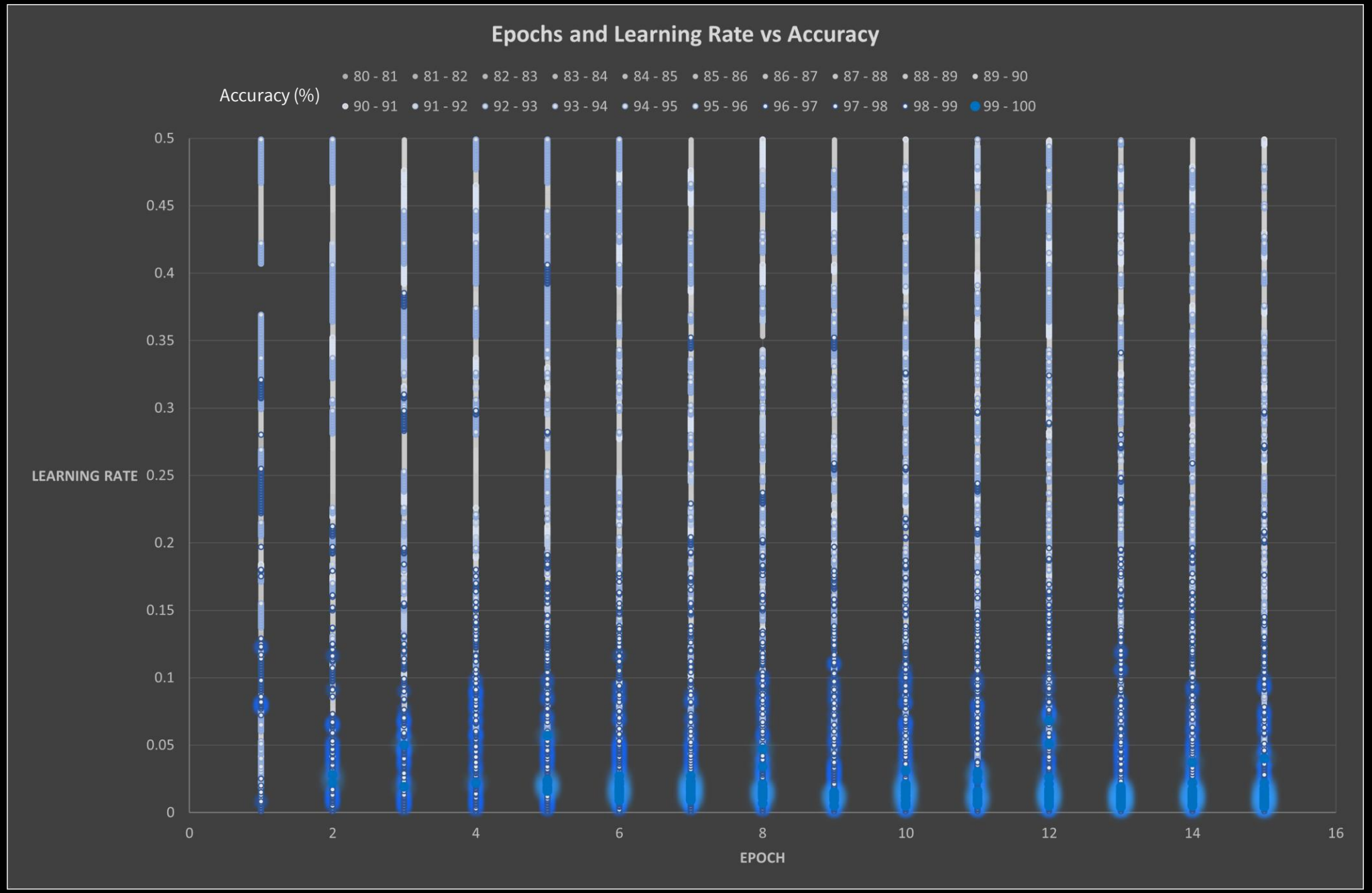
Model optimization

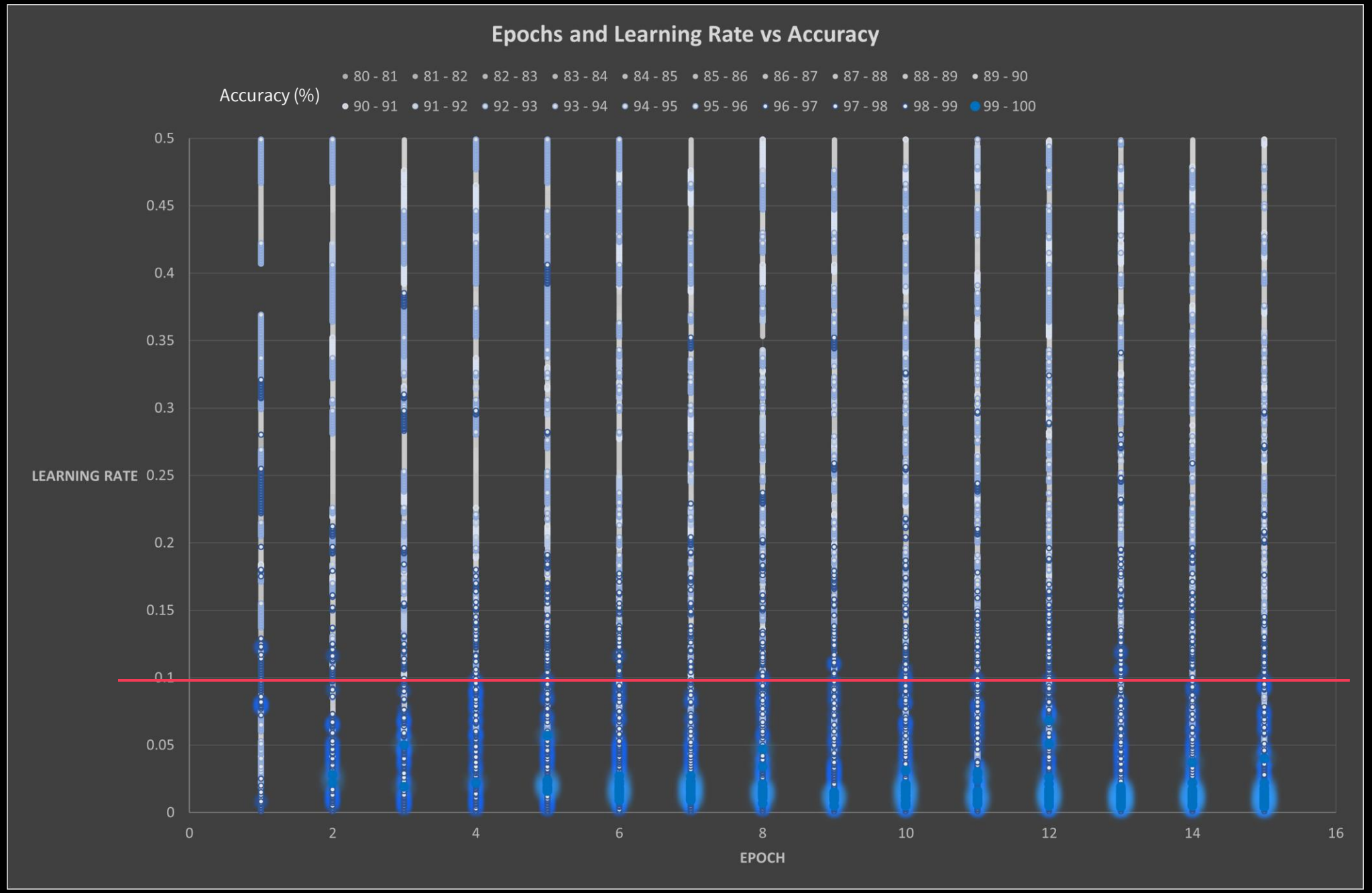
Procedure:

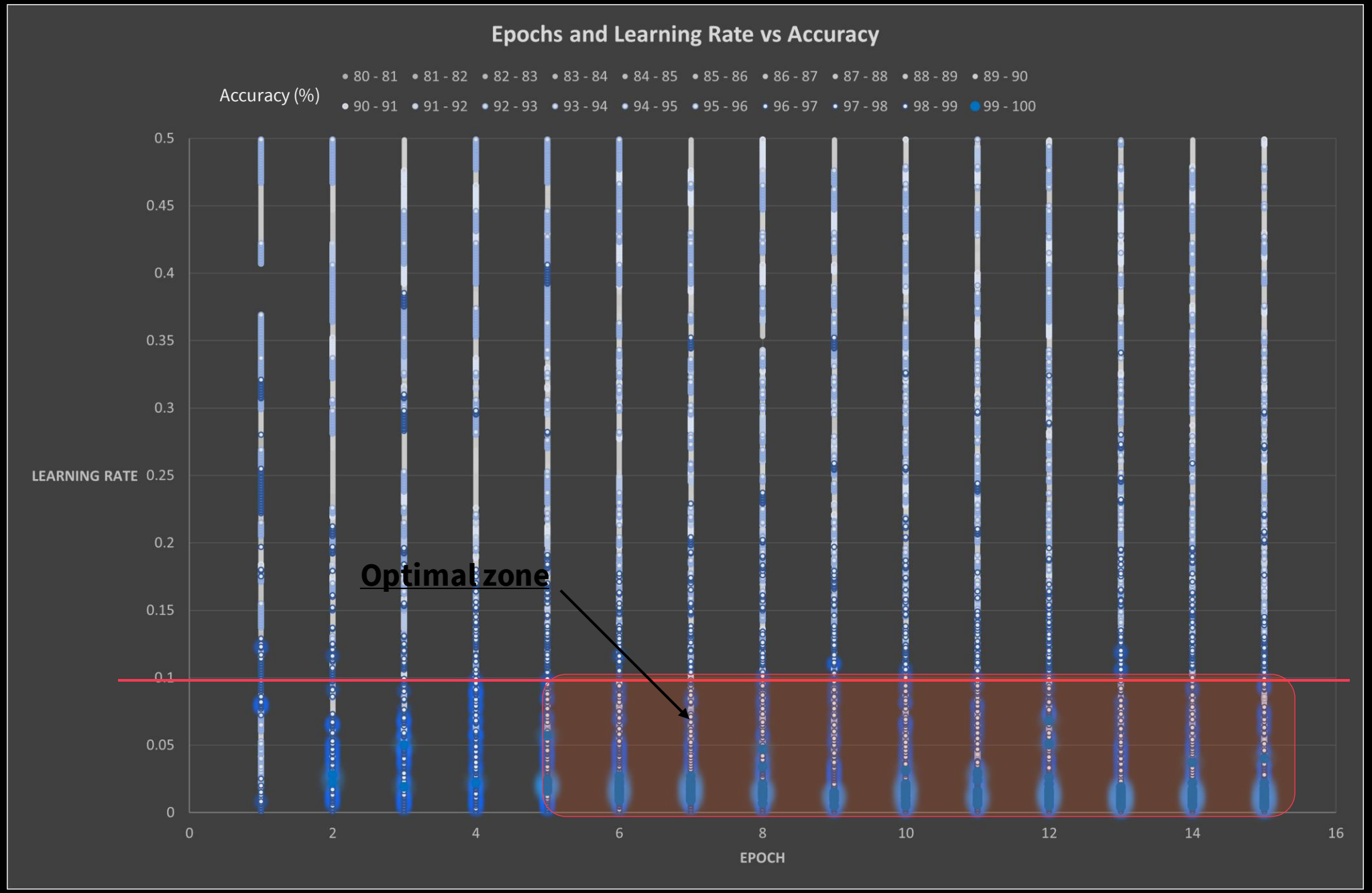
- Run through range of epochs and learning rate
- Run the model
- Find the point of maximum accuracy

- Most useful for non-linearly separable data, as linearly separable data easily reaches 100% accuracy...


Accuracy vs Learning Rate
Epochs = 10
100
95
90
85
80
75
ACCURACY (%)
0.001
0.012
0.023
0.034
0.045
0.056
0.067
0.078
0.089
0.1
0.111
0.122
0.133
0.144
0.155
0.166
0.177
0.188
0.199
0.21
0.221
0.232
0.243
LEARNING RATE


Epochs and Learning Rate vs Accuracy
Accuracy (%)
80 - 81
81 - 82
82 - 83
83 - 84
84 - 85
85 - 86
86 - 87
87 - 88
88 - 89
89 - 90
90 - 91
91 - 92
92 - 93
93 - 94
94 - 95
95 - 96
96 - 97
97 - 98
98 - 99
99 - 100
LEARNING RATE
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
2
4
6
8
10
12
14
16
EPOCH


Epochs and Learning Rate vs Accuracy
Accuracy (%)
80 - 81
81 - 82
82 - 83
83 - 84
84 - 85
85 - 86
86 - 87
87 - 88
88 - 89
89 - 90
90 - 91
91 - 92
92 - 93
93 - 94
94 - 95
95 - 96
96 - 97
97 - 98
98 - 99
99 - 100
LEARNING RATE
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
2
4
6
8
10
12
14
16
EPOCH


Epochs and Learning Rate vs Accuracy
Accuracy (%)
80 - 81
81 - 82
82 - 83
83 - 84
84 - 85
85 - 86
86 - 87
87 - 88
88 - 89
89 - 90
90 - 91
91 - 92
92 - 93
93 - 94
94 - 95
95 - 96
96 - 97
97 - 98
98 - 99
99 - 100
LEARNING RATE
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
2
4
6
8
10
12
14
16
EPOCH
Optimal zone

Iris Database

- Fixed learning rate = 0.02
  - reached 100% accuracy in 5 epochs
- Optimized learning rate = 0.05
  - reached 100% accuracy in 1 epoch

Balloons Database

- Fixed learning rate = 0.02
  - reached 100% accuracy in 5 epochs
- Optimized learning rate = 0.34
  - reached 100% accuracy in 1 epoch

Banknote Database

- Fixed learning rate = 0.02
  - reached 99.27% accuracy in 10 epochs
- Optimized learning rate = 0.02
  - reached 99.27% accuracy in 1 epoch

Computation costs

- Measured in epochs (O(N^3 + N) each iteration)

- Efficient for linearly separable data (~10 epochs needed)
- Less efficient for non-linearly separable data
  - How many epochs do we run?

Disadvantages

- In order to reach 100% accuracy, data must be linearly separable
- In the real world, linearly separable data is *rare*
- As number of attributes increase, the data may become less linearly separable
- Irrelevant attributes (outliers):
  - Datasets with at least 1 outlier can never be linearly separable)

Thank you! :)