


# INTRO TO MACHINE LEARNING

The background of the image is a dark, industrial scene filled with several Terminator robots. They are metallic, with glowing red eyes and intricate mechanical details. The robots are positioned in various ways, some standing and others in motion, creating a sense of a large-scale invasion or battle. The lighting is dim, with some highlights on the robots' surfaces and their glowing eyes.

**(OR: WHY SKYNET  
WON'T KILL US  
ANYTIME SOON)**

**WHAT?**

A COMPUTER PROGRAM IS SAID  
TO LEARN FROM EXPERIENCE  $E$   
WITH RESPECT TO SOME CLASS  
OF TASKS  $T$  AND PERFORMANCE  
MEASURE  $P$ . IF IT'S  
PERFORMANCE AT TASKS IN  $T$ ,  
AS MEASURED BY  $P$ , IMPROVES  
WITH EXPERIENCE  $E$

**A PROGRAM THAT  
IMPROVES ITSELF  
WITHOUT OUR  
ACTIVE INPUT.**

WHY?

# **SUPERVISED VS. UNSUPERVISED LEARNING**

# **SUPERVISED LEARNING: LINEAR REGRESSION VS. CLASSIFICATION**



# UNSUPERVISED LEARNING

HOW?

# FEATURES

# TRAINING SET

$$(x^{(i)}, y^{(i)})$$

# HYPOTHESIS FUNCTION

$$h_{\Theta}(x)$$

# COST FUNCTION

$$J(\Theta)$$

# OPTIMISATION FUNCTION

**LET'S ML!**



# LOGISTIC REGRESSION

$$h_{\Theta}(x) = \frac{1}{1 + e^{-\Theta^T x}}$$

$$h_{\Theta}(x) = P(y = 1|x; \Theta)$$

$$J(\Theta) = \frac{1}{m} \sum_{i=1}^m \textit{Cost}(h_{\Theta}(x), y)$$

$$\begin{aligned} \textit{Cost}(h_{\Theta}(x), y) = & -y \cdot \log(h_{\Theta}(x)) - \\ & (1 - y) \cdot (\log(1 - h_{\Theta}(x))) \end{aligned}$$

$$y = 1$$

$$Cost(h_{\Theta}(x), y) = -y \cdot \log(h_{\Theta}(x)) - (1 - y) \cdot (\log(1 - h_{\Theta}(x)))$$

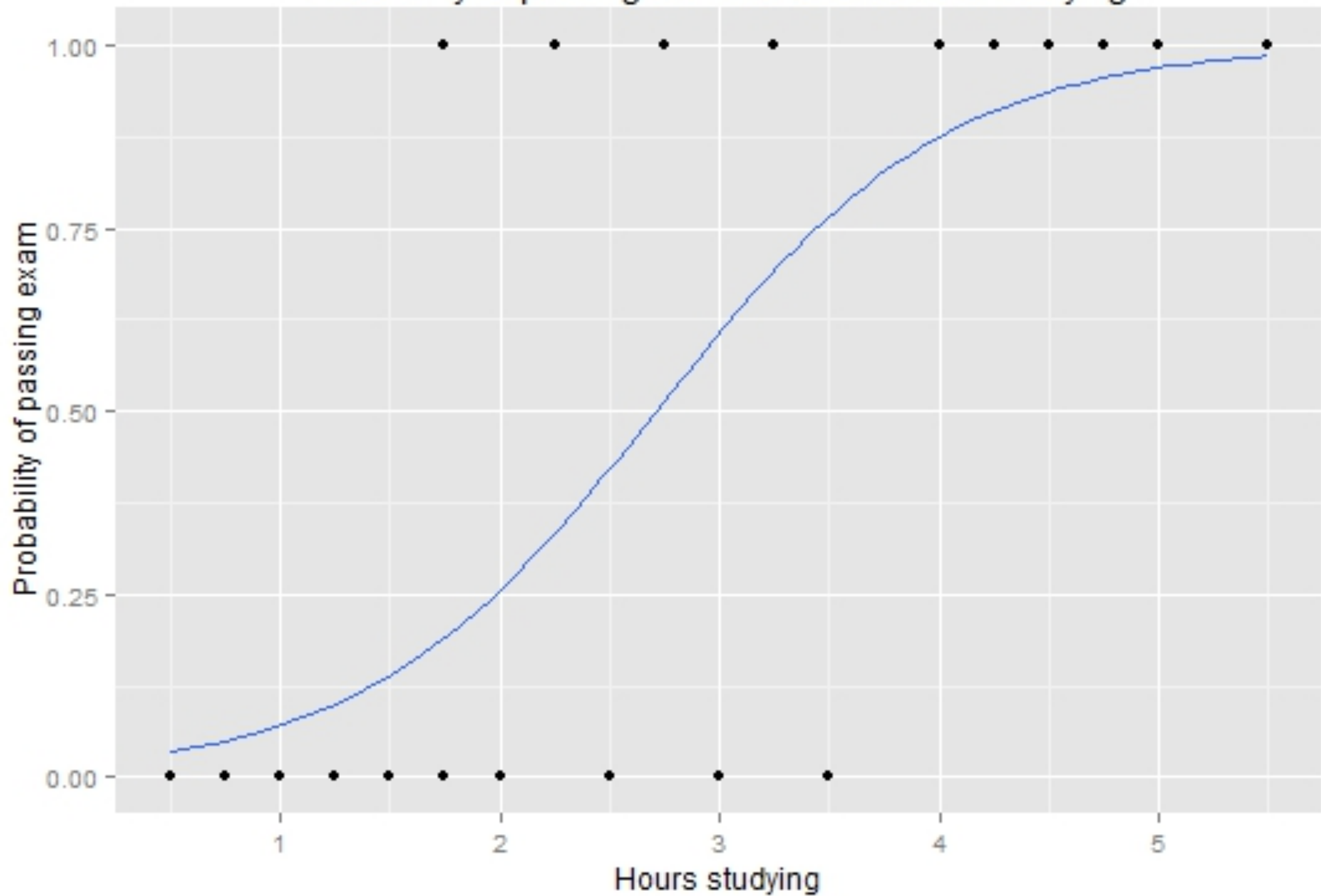
$$Cost(h_{\Theta}(x), y) = -y \cdot \log(h_{\Theta}(x))$$

$$y = 0$$

$$Cost(h_{\Theta}(x), y) = -y \cdot \log(h_{\Theta}(x)) - (1 - y) \cdot (\log(1 - h_{\Theta}(x)))$$

$$Cost(h_{\Theta}(x), y) = -\log(1 - h_{\Theta}(x))$$

Probability of passing exam versus hours of studying

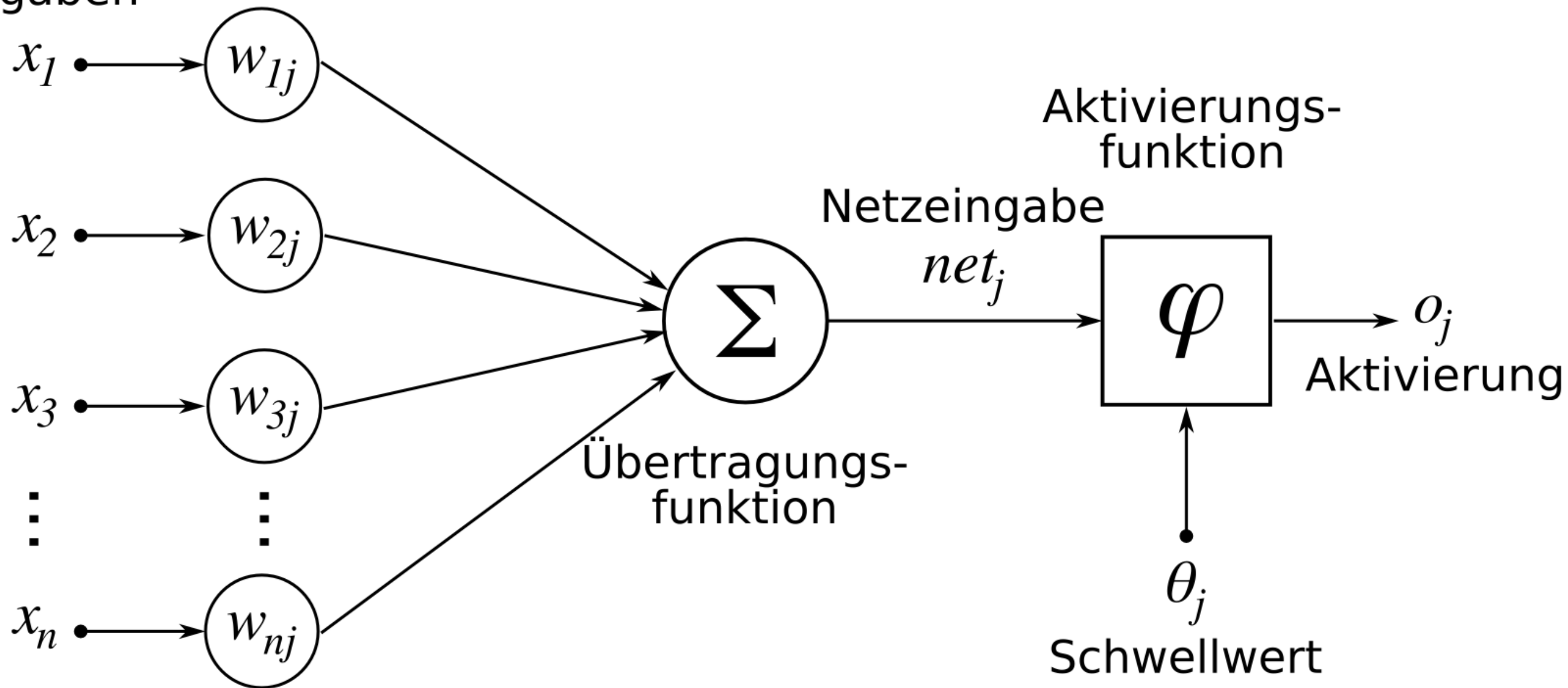


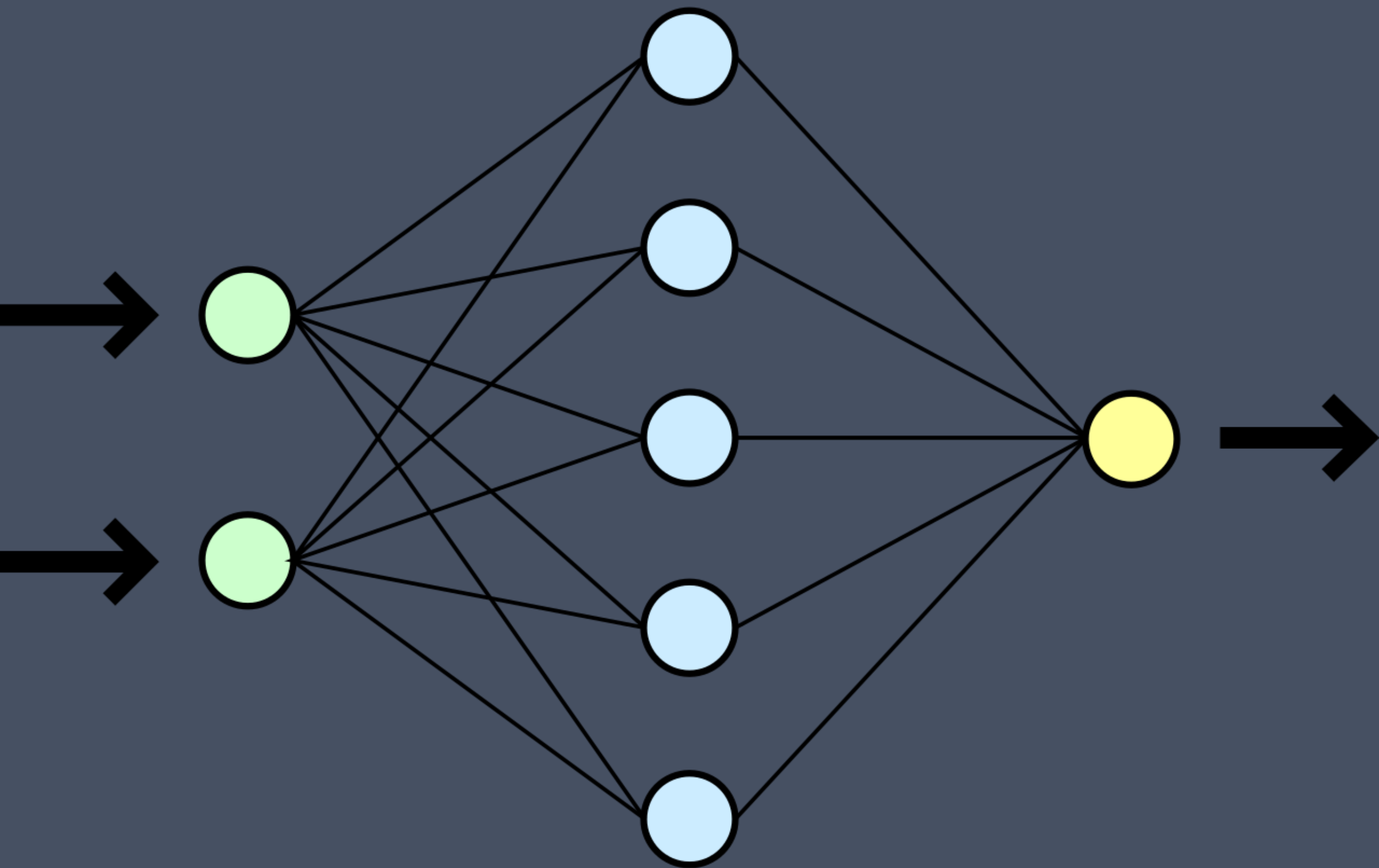


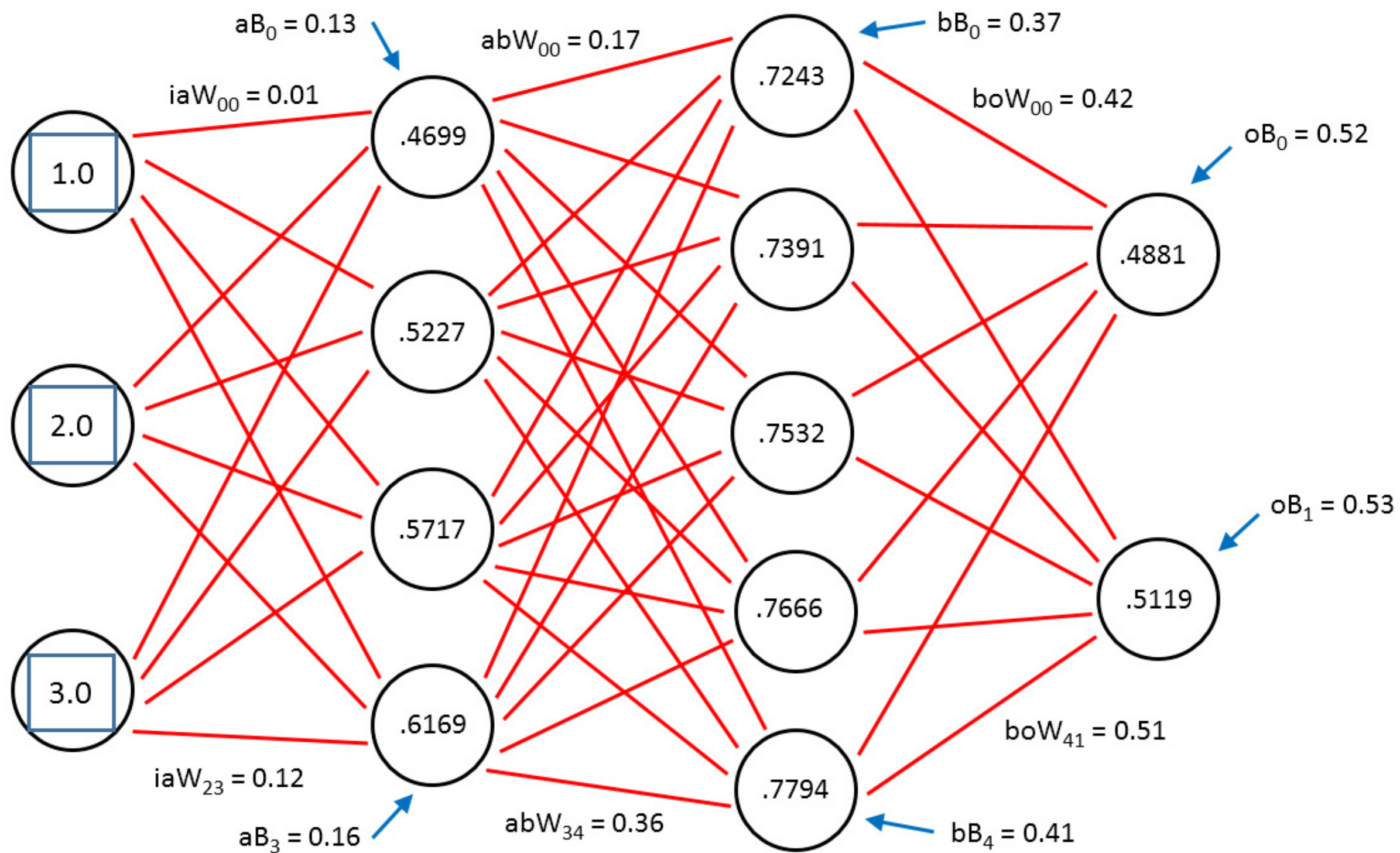
# MULTICLASS CLASSIFICATION

# NEURAL NETWORKS

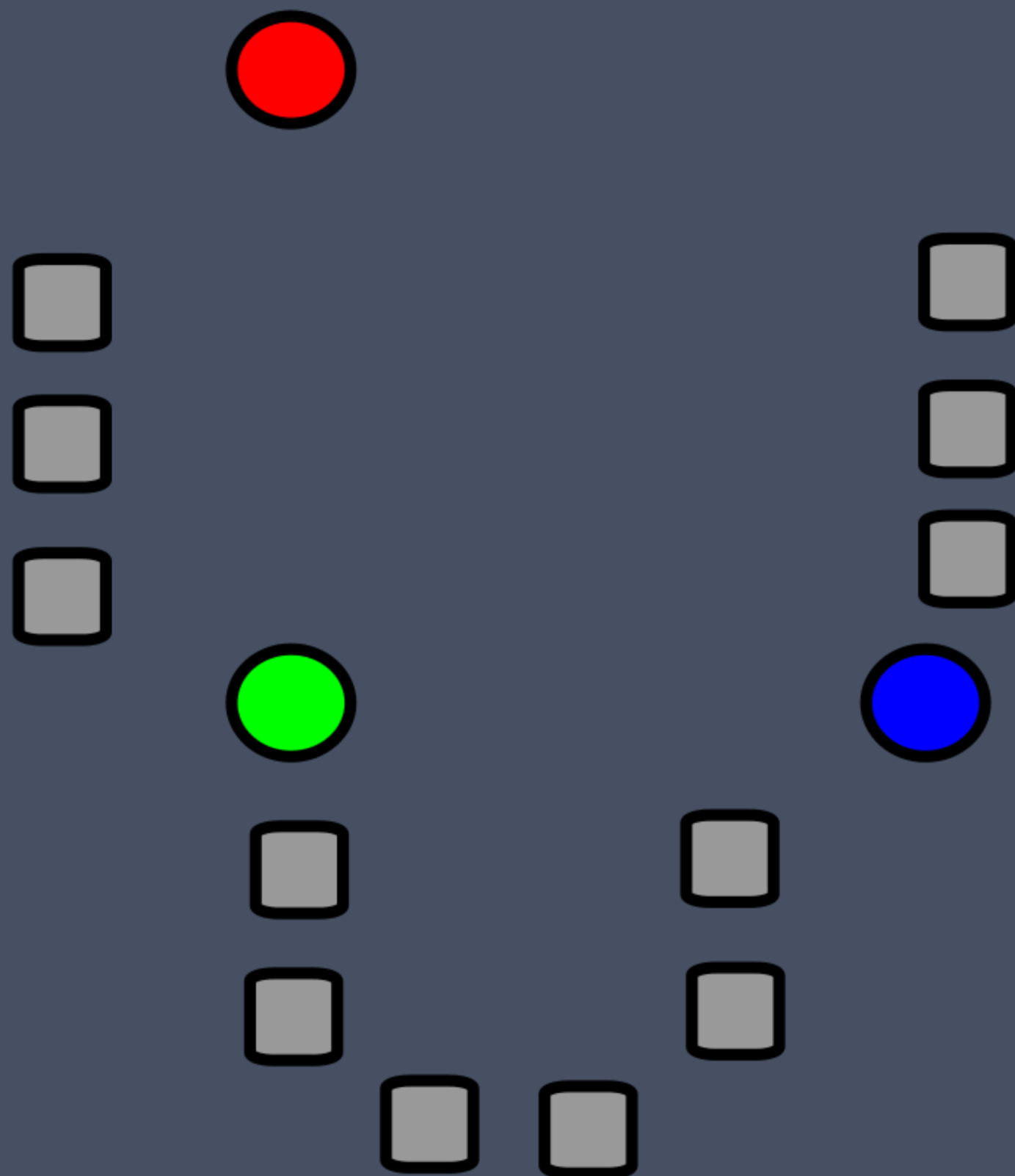
Eingaben Gewichtungen

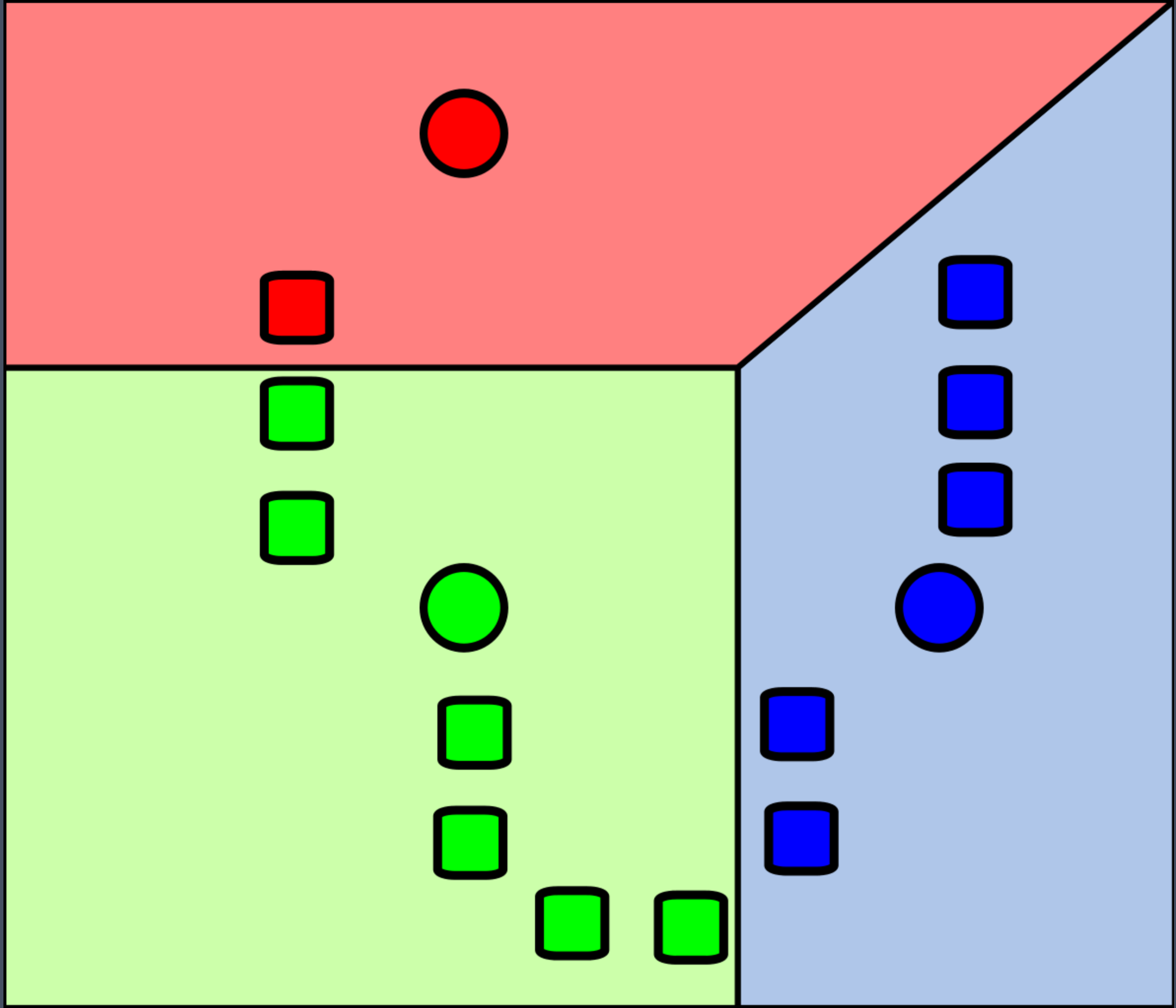




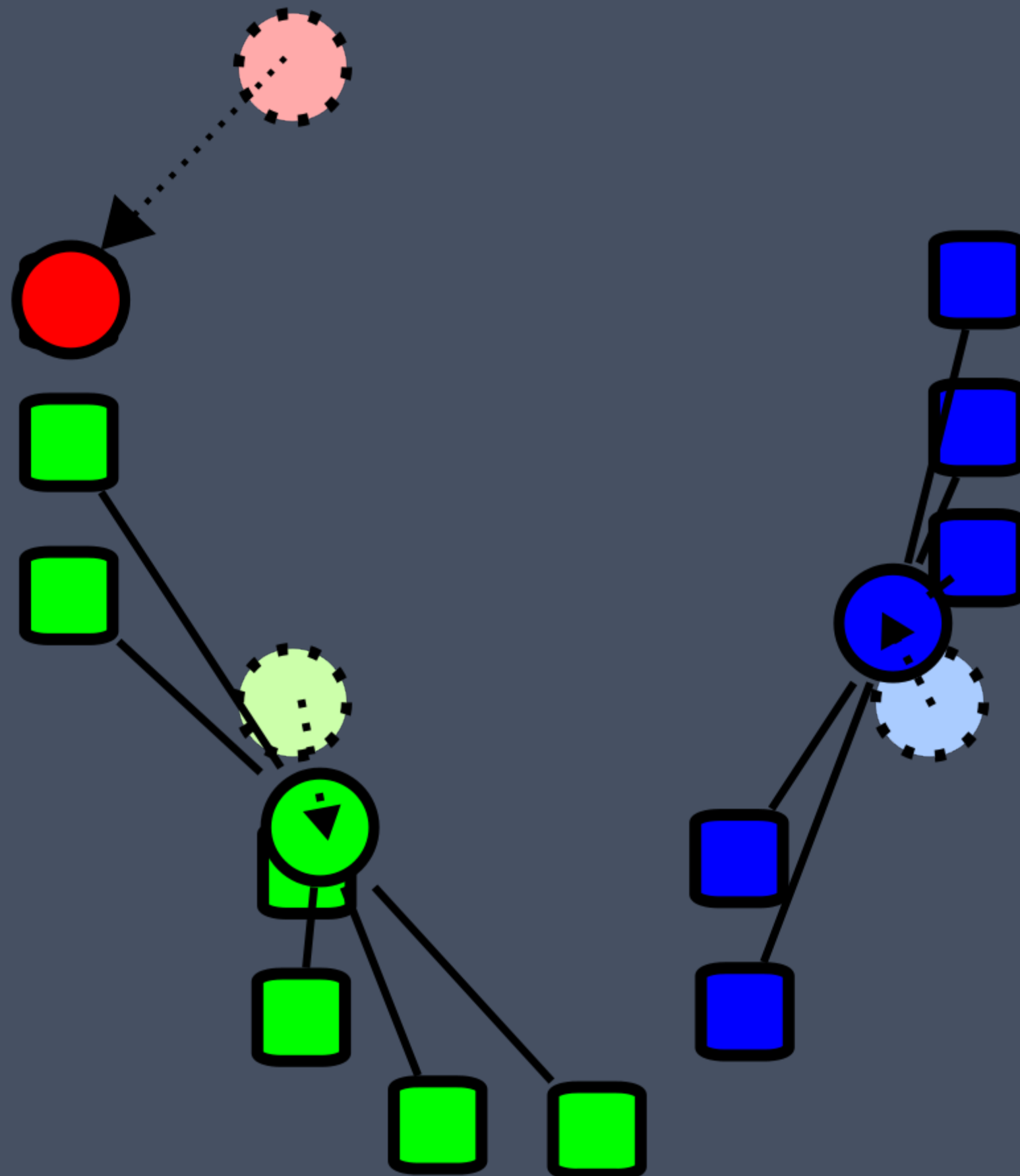


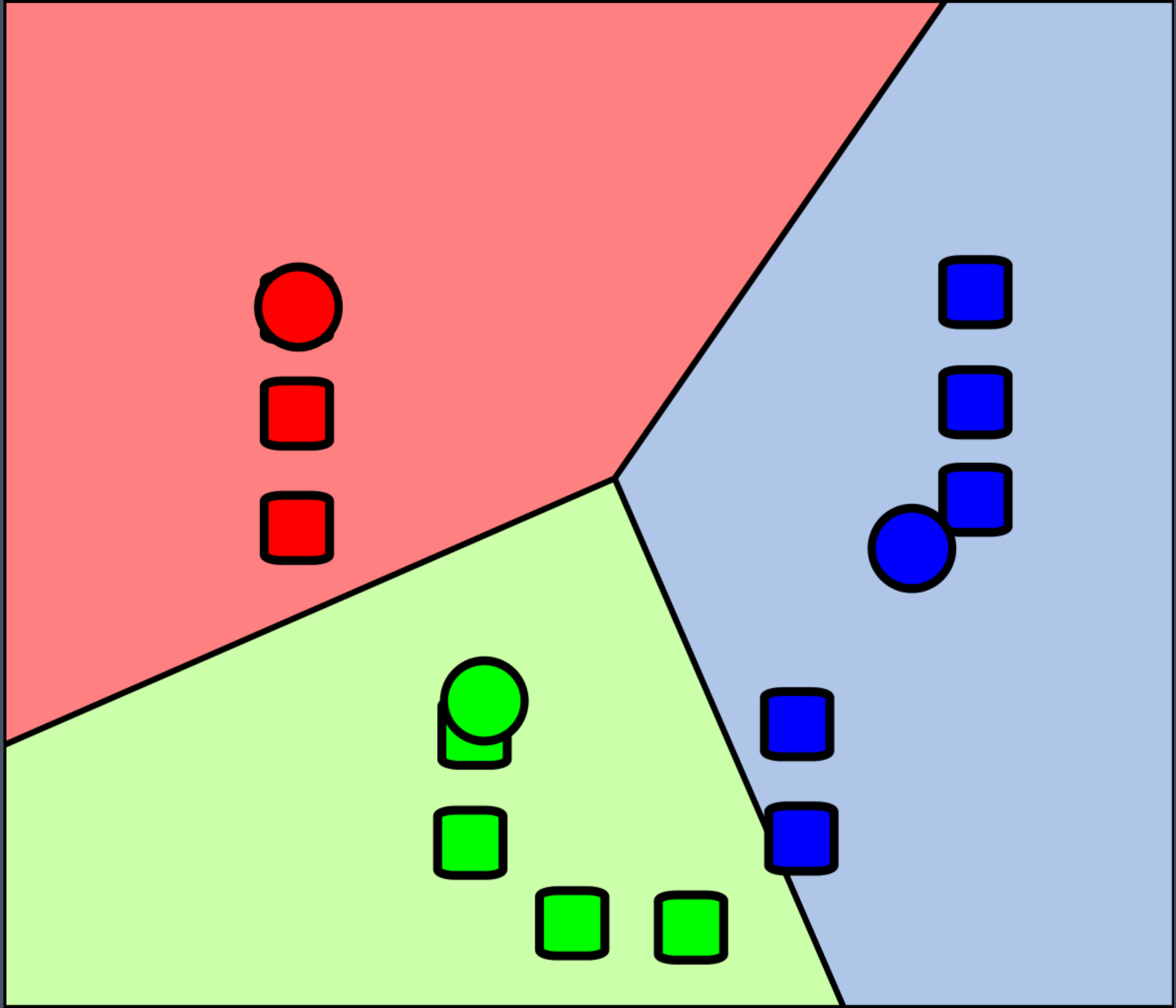
# K-MEANS











# SUPPORT VECTOR MACHINES ANOMALY DETECTION

# POSSIBLE PROBLEMS

**OVERFITTING**  
**(HIGH VARIANCE)**

# UNDERFITTING

## (HIGH BIAS)

ML'S DARK  
SECRET

90% OF AN ML PROJECT  
HAS NOTHING  
TO DO WITH ML



ML IS RESTRICTED  
TO VERY SPECIFIC  
PROBLEMS

**ML IN RUBY**

# Ruby

---

## Natural Language Processing

- [Treat](#) - Text REtrieval and Annotation Toolkit, definitely the most comprehensive toolkit I've encountered so far for Ruby
- [Ruby Linguistics](#) - Linguistics is a framework for building linguistic utilities for Ruby objects in any language. It includes a generic language-independent front end, a module for mapping language codes into language names, and a module which contains various English-language utilities.
- [Stemmer](#) - Expose libstemmer\_c to Ruby
- [Ruby Wordnet](#) - This library is a Ruby interface to WordNet
- [Raspel](#) - raspell is an interface binding for ruby
- [UEA Stemmer](#) - Ruby port of UEALite Stemmer - a conservative stemmer for search and indexing
- [Twitter-text-rb](#) - A library that does auto linking and extraction of usernames, lists and hashtags in tweets

## General-Purpose Machine Learning

- [Machine Learning](#) - The Machine Learning examples in Ruby
- [Machine Learning](#) - Ruby
- [Ruby Mahout](#) - JFRuby Mahout is a wrapper that unleashes the power of Apache Mahout into the world of Ruby.
- [CardMagic-Classification](#) - A general classification module for ResNet type classification
- [rb-libsvm](#) - Ruby language bindings for LIBSVM which is a Library for Support Vector Machines

## Data Analysis / Data Visualization

- [rsruby](#) - Ruby - R bridge
- [data-visualization-ruby](#) - Source code and supporting content for my Ruby Manor presentation on Data Visualisation with Ruby
- [ruby-plot](#) - gnuplot wrapper for ruby, especially for plotting roc curves into svg files
- [plot-rb](#) - A plotting library in Ruby built on top of Vega and D3.
- [scruffy](#) - A beautiful graphing toolkit for Ruby
- [SciRuby](#)
- [Glean](#) - A data management tool for humans
- [Bioruby](#)
- [Arel](#)

## Misc

- [Big Data For Chimps](#)
- [Listof](#) - Community based data collection, packed in gem. Get list of pretty much anything (stop words, countries, non words) in txt, json or hash. [Demo/Search for a list](#)

FEDERAL

GALAXY

TOP NEWS

ENLIST

EXIT

> COURSERA ML COURSE (STARTS MAY 16!)

> THIS QUORA ANSWER

> KAGGLE

> 'PROGRAMMING COLLECTIVE INTELLIGENCE' BY  
O'REILLY

> AWESOME MACHINE LEARNING

✈ WOULD YOU LIKE TO KNOW MORE?

**THE END**

# PICTURE CREDITS

- > [HTTPS://EN.WIKIPEDIA.ORG/WIKI/  
FILE:EXAMPASSLOGISTIC\\_CURVE.JPEG](https://en.wikipedia.org/wiki/File:EXAMPASSLOGISTIC_CURVE.JPEG)
- > [HTTPS://DE.WIKIPEDIA.ORG/WIKI/  
DATEI:ARTIFICIALNEURONMODEL\\_DEUTSCH.PNG](https://de.wikipedia.org/wiki/Datei:ArtificialNeuronModel_Deutsch.PNG)
- > [HTTPS://DE.WIKIPEDIA.ORG/WIKI/  
DATEI:NEURAL\\_NETWORK.SVG](https://de.wikipedia.org/wiki/Datei:Neural_Network.SVG)

# PICTURE CREDITS

- > [HTTPS://COMMONS.WIKIMEDIA.ORG/WIKI/  
FILE:KMEANSEXAMPLESTEP1.SVG](https://commons.wikimedia.org/wiki/File:KMEANSEXAMPLESTEP1.SVG)
- > [HTTPS://COMMONS.WIKIMEDIA.ORG/WIKI/  
FILE:KMEANSEXAMPLESTEP2.SVG](https://commons.wikimedia.org/wiki/File:KMEANSEXAMPLESTEP2.SVG)
- > [HTTPS://COMMONS.WIKIMEDIA.ORG/WIKI/  
FILE:KMEANSEXAMPLESTEP3.SVG](https://commons.wikimedia.org/wiki/File:KMEANSEXAMPLESTEP3.SVG)
- > [HTTPS://COMMONS.WIKIMEDIA.ORG/WIKI/  
FILE:KMEANSEXAMPLESTEP4.SVG](https://commons.wikimedia.org/wiki/File:KMEANSEXAMPLESTEP4.SVG)