Removing the blindfold

Visualising statistical models

Hadley Wickham
Assistant Professor
Dobelman Family Junior Chair
Department of Statistics
Rice University

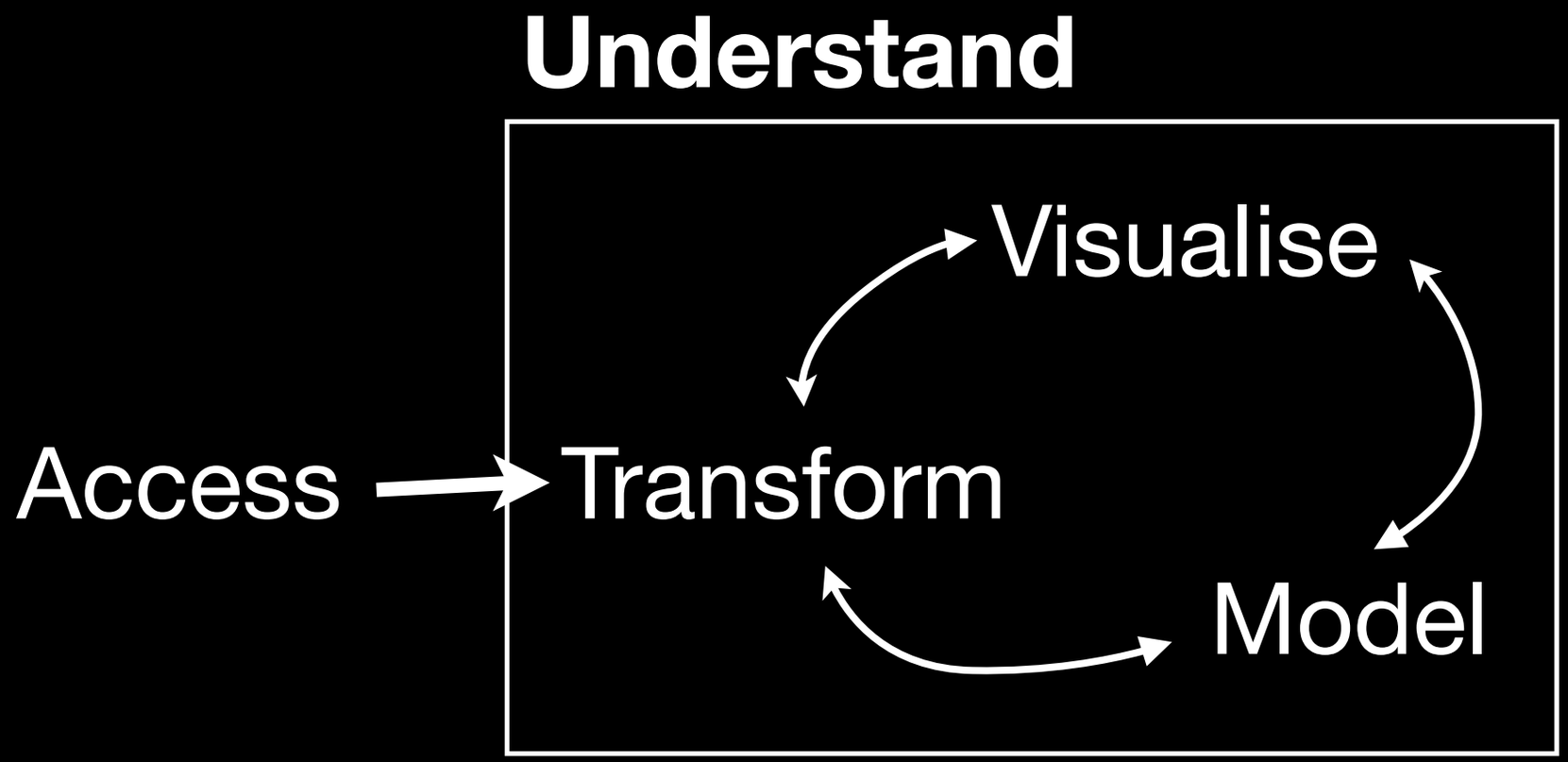
Why?

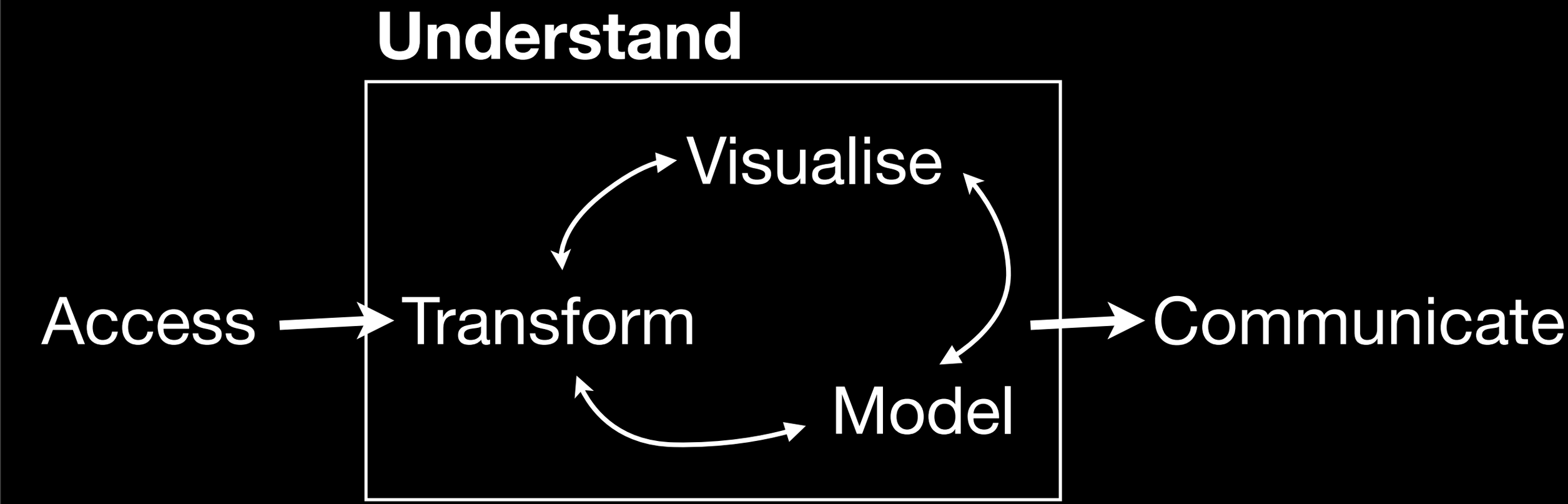
Friday, 10 September 2010

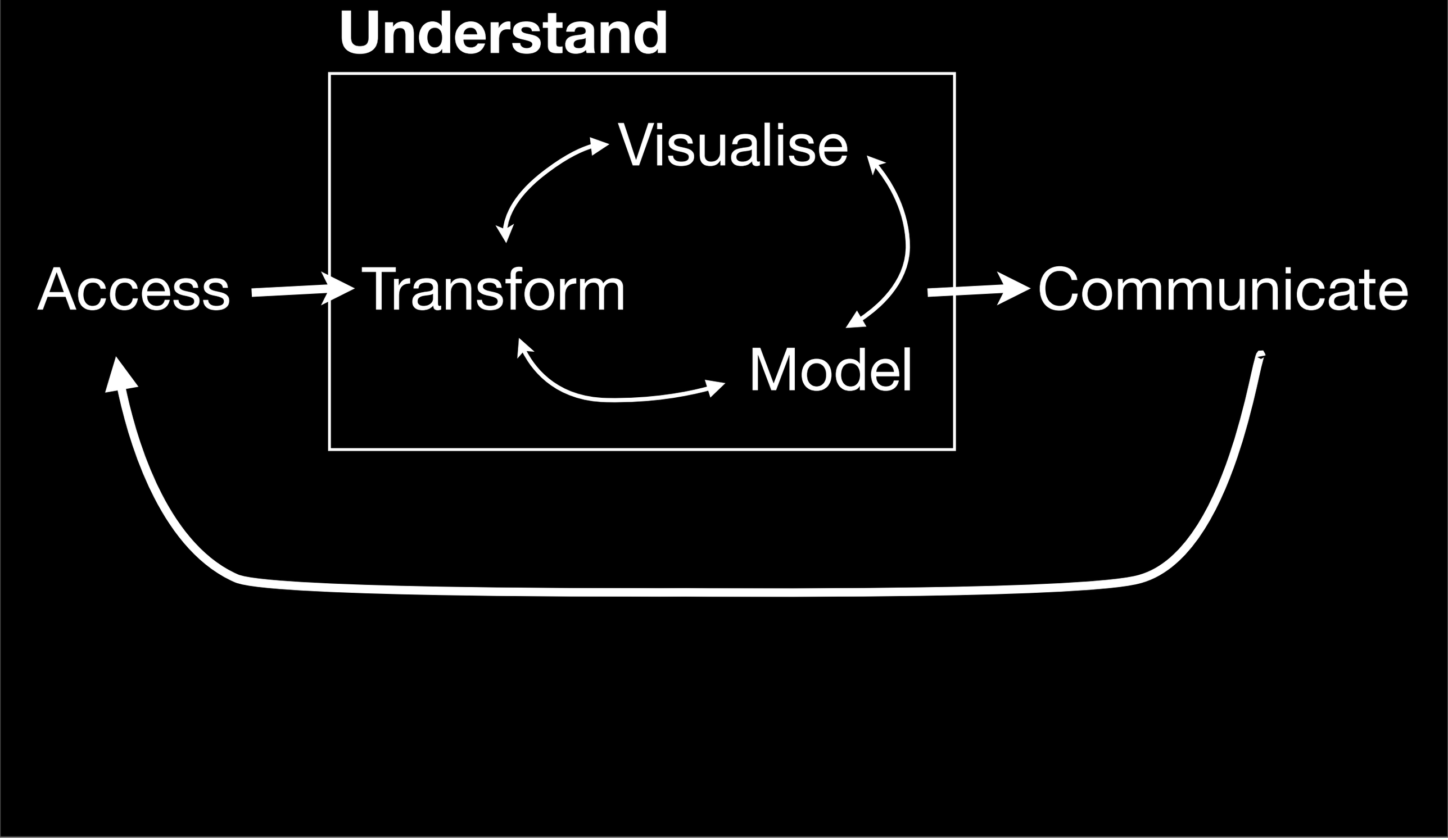
Access

Understand

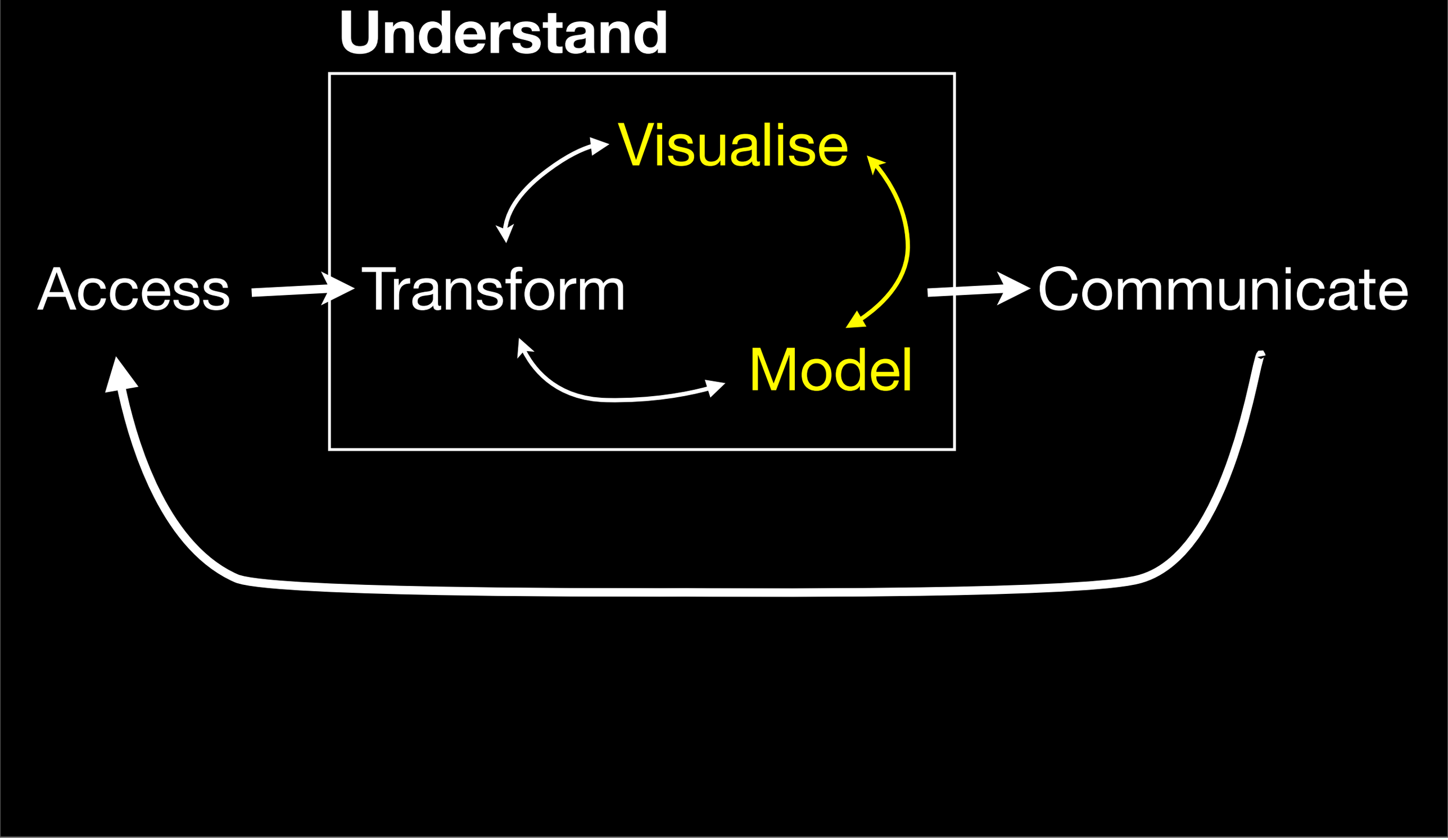
Access →


Understand
Visualise
Access
Transform
Model


Understand
Visualise
Access
Transform
Communicate
Model


Understand
Visualise
Access
Transform
Communicate
Model

Visualisation	Model
+ Uncovers the unexpected - Slow - Cognitive biases	+ Mathematically well founded + Fast - Only discovers what we anticipate


Understand
Visualise
Access
Transform
Communicate
Model

Neural networks

Display the model in the data space

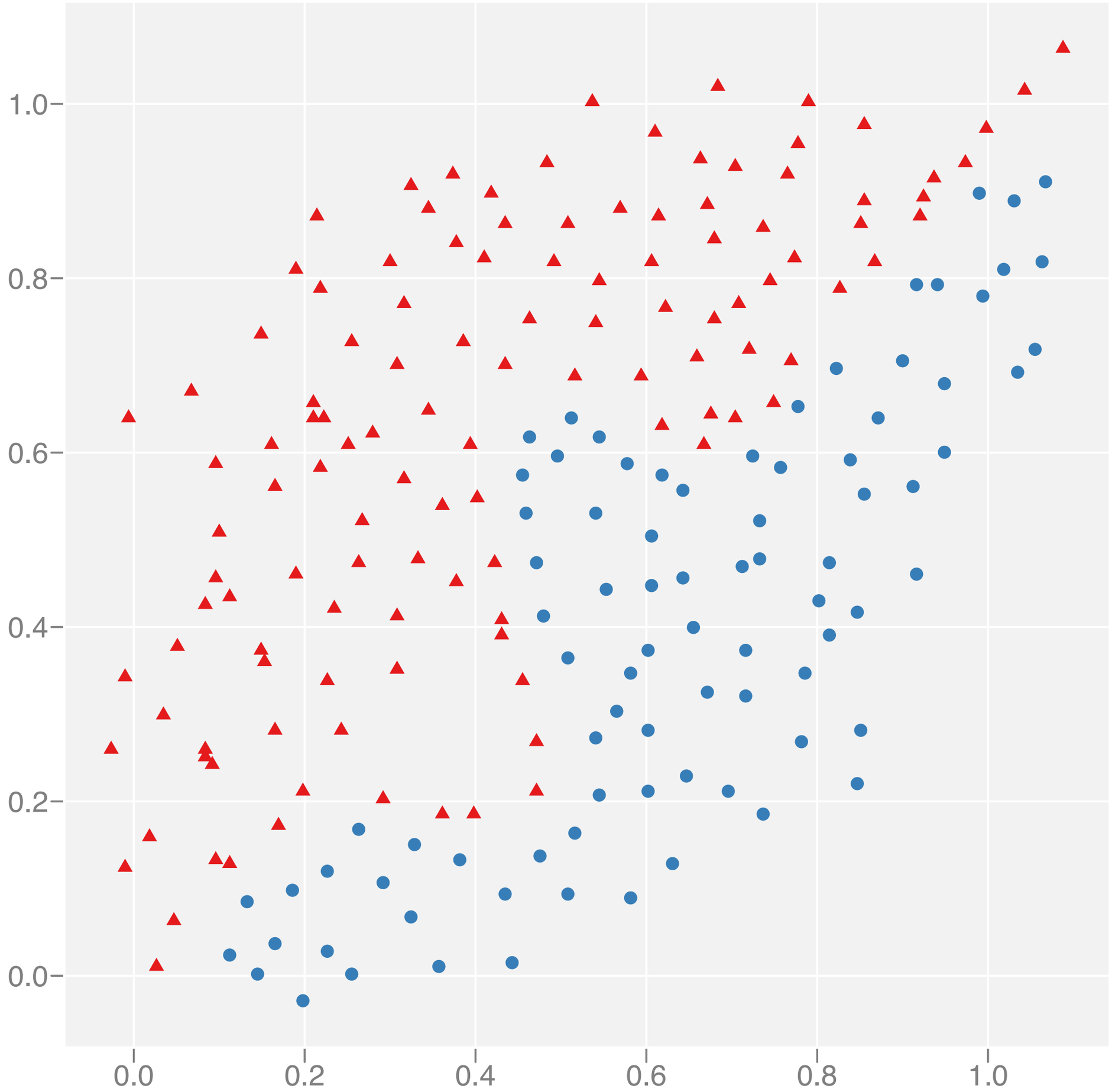
Look at many members of a collection

Explore the process of fitting, not just the end result

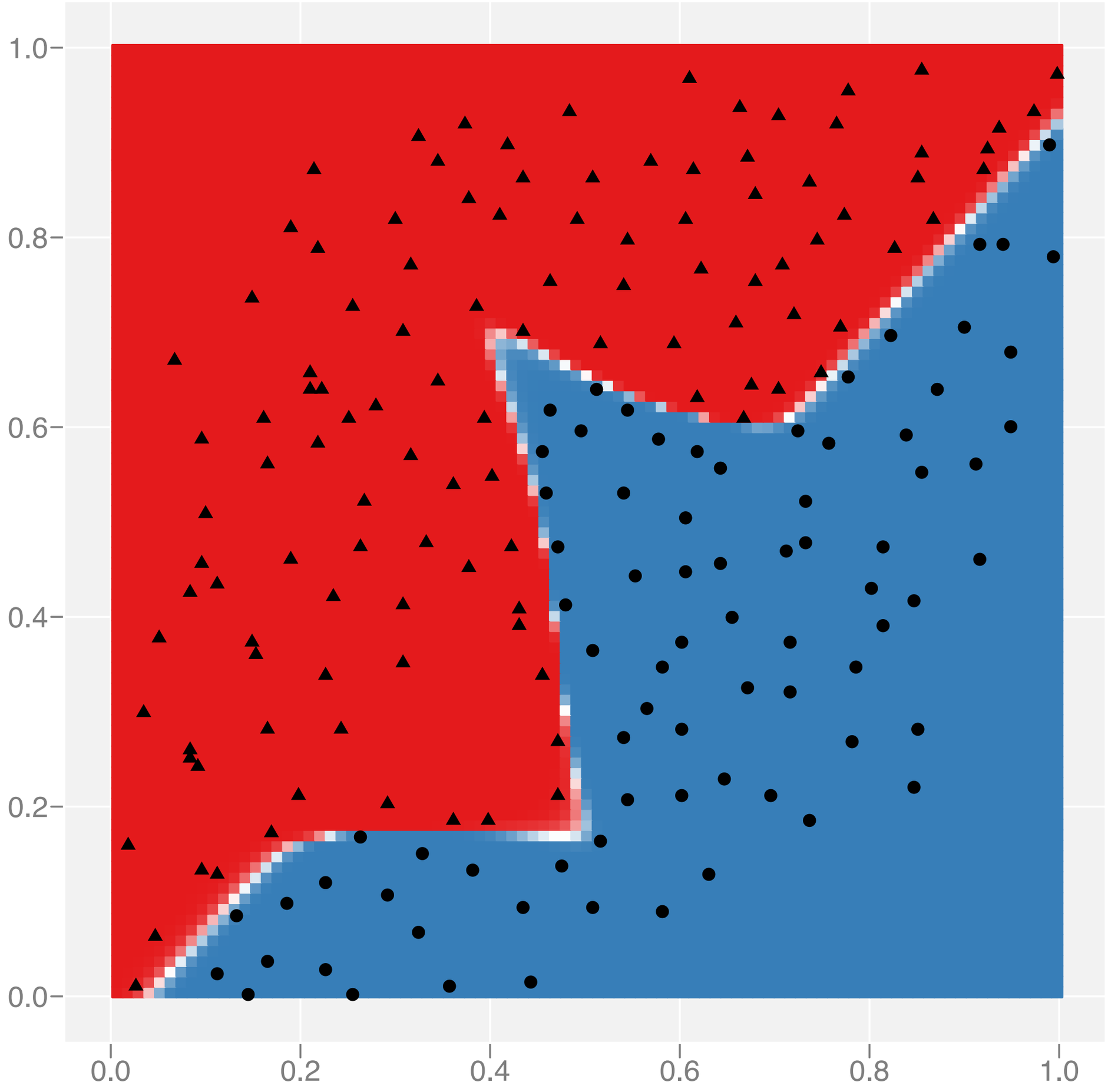
Neural networks

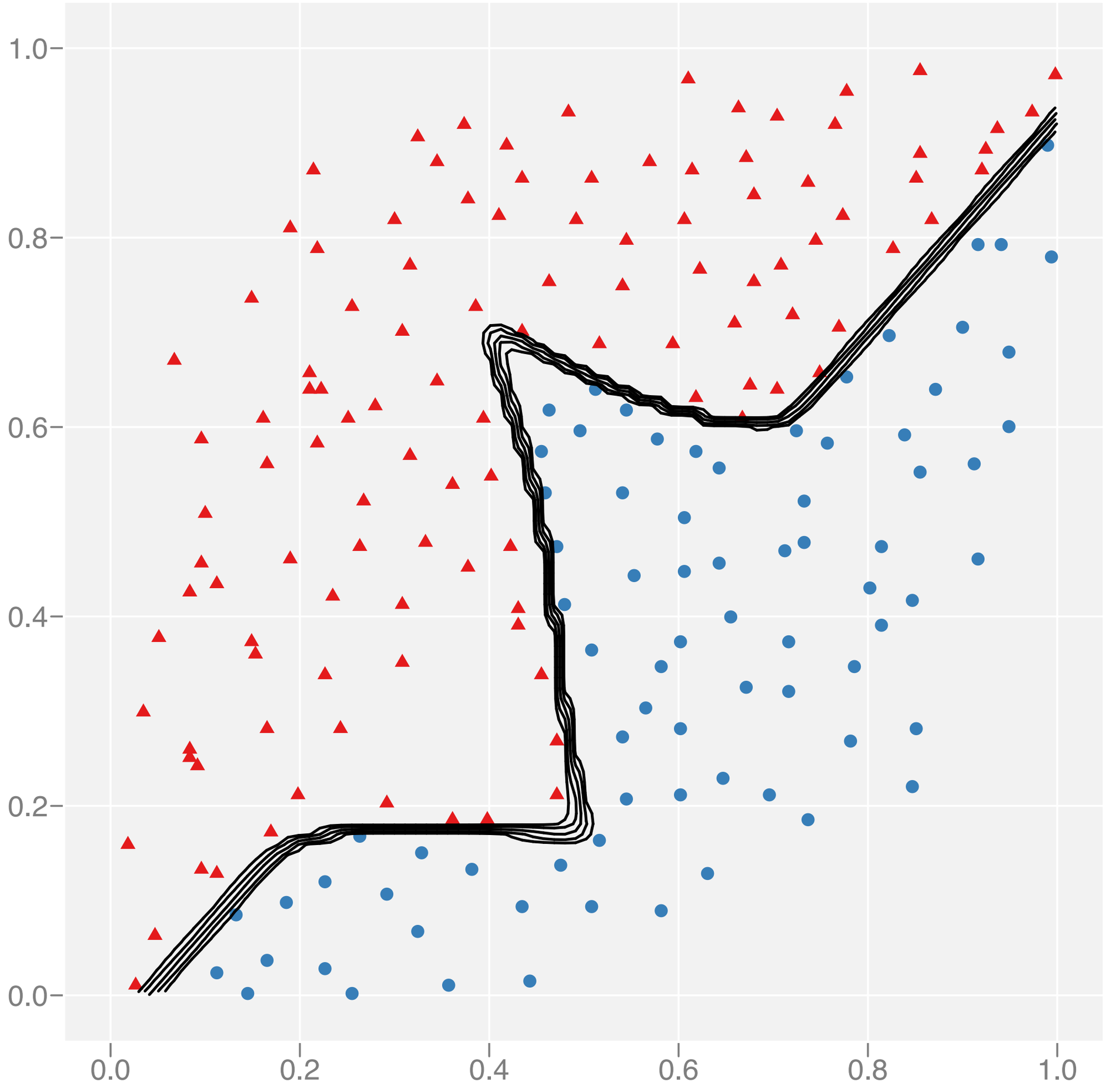
- Modelled on the way that brains work
- Normally treated as a black box. Can we gain more insight into how they work?

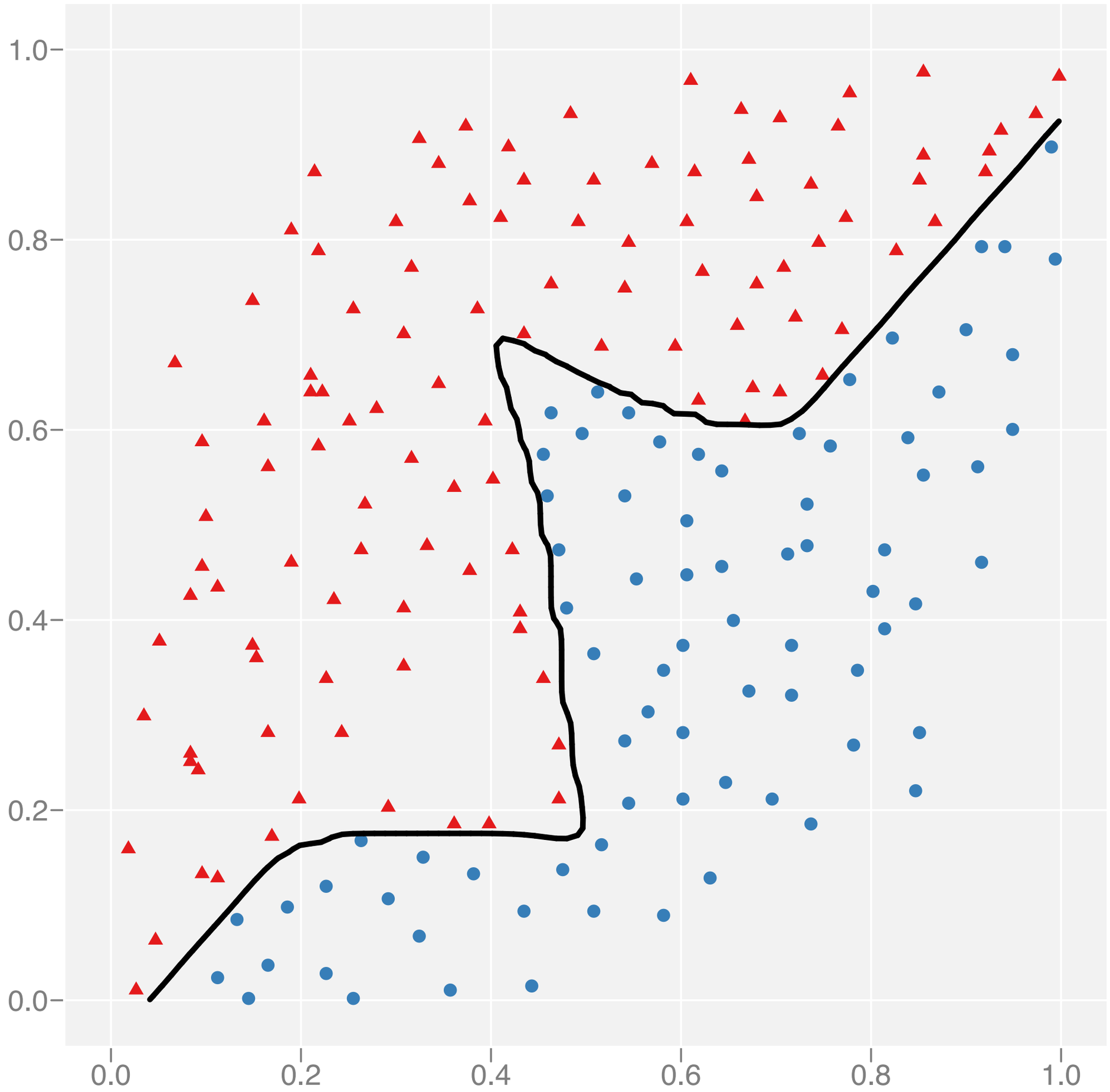
- Single hidden-layer neural network: `nnet` R package


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

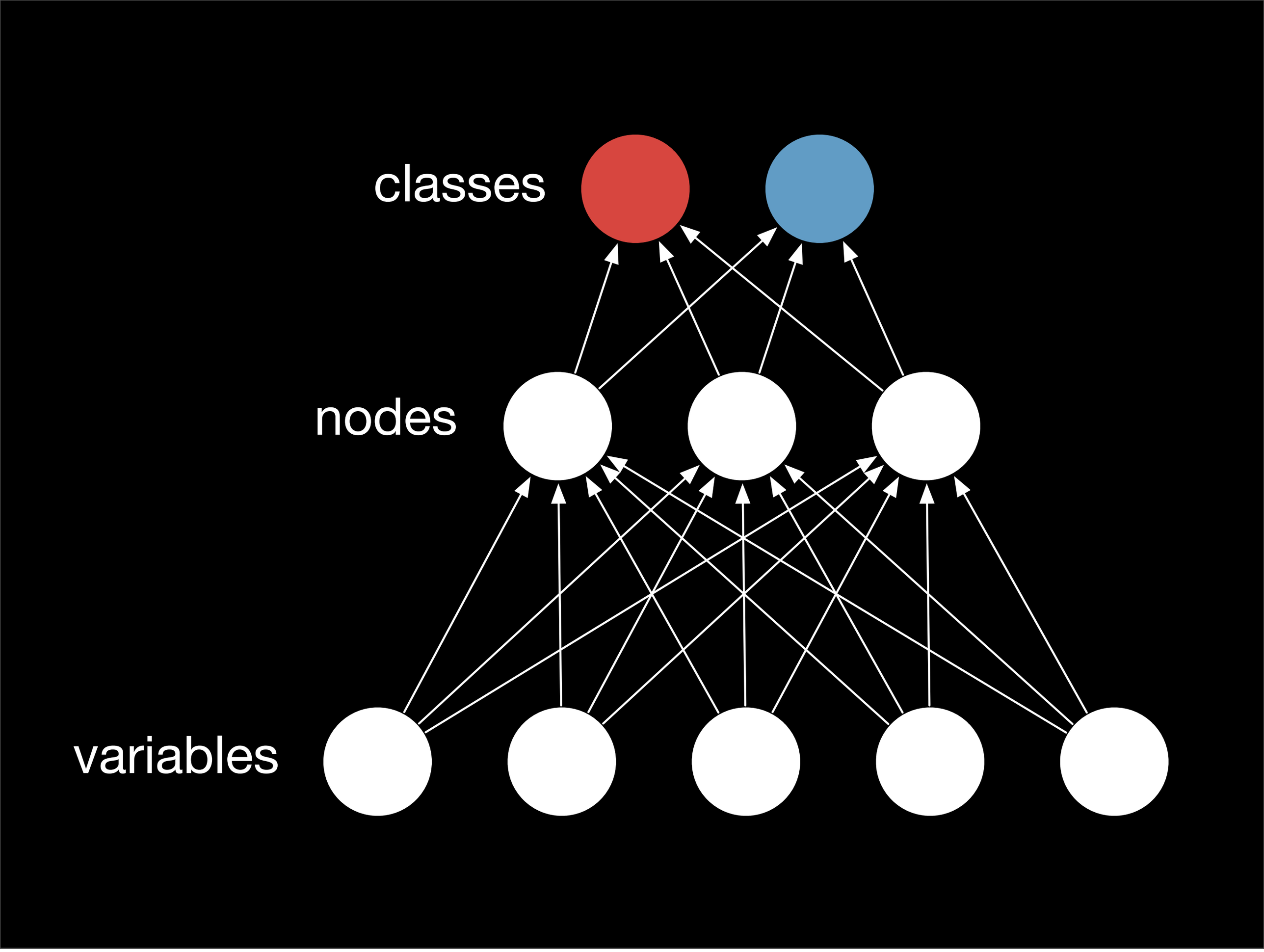
Display the model in data space


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

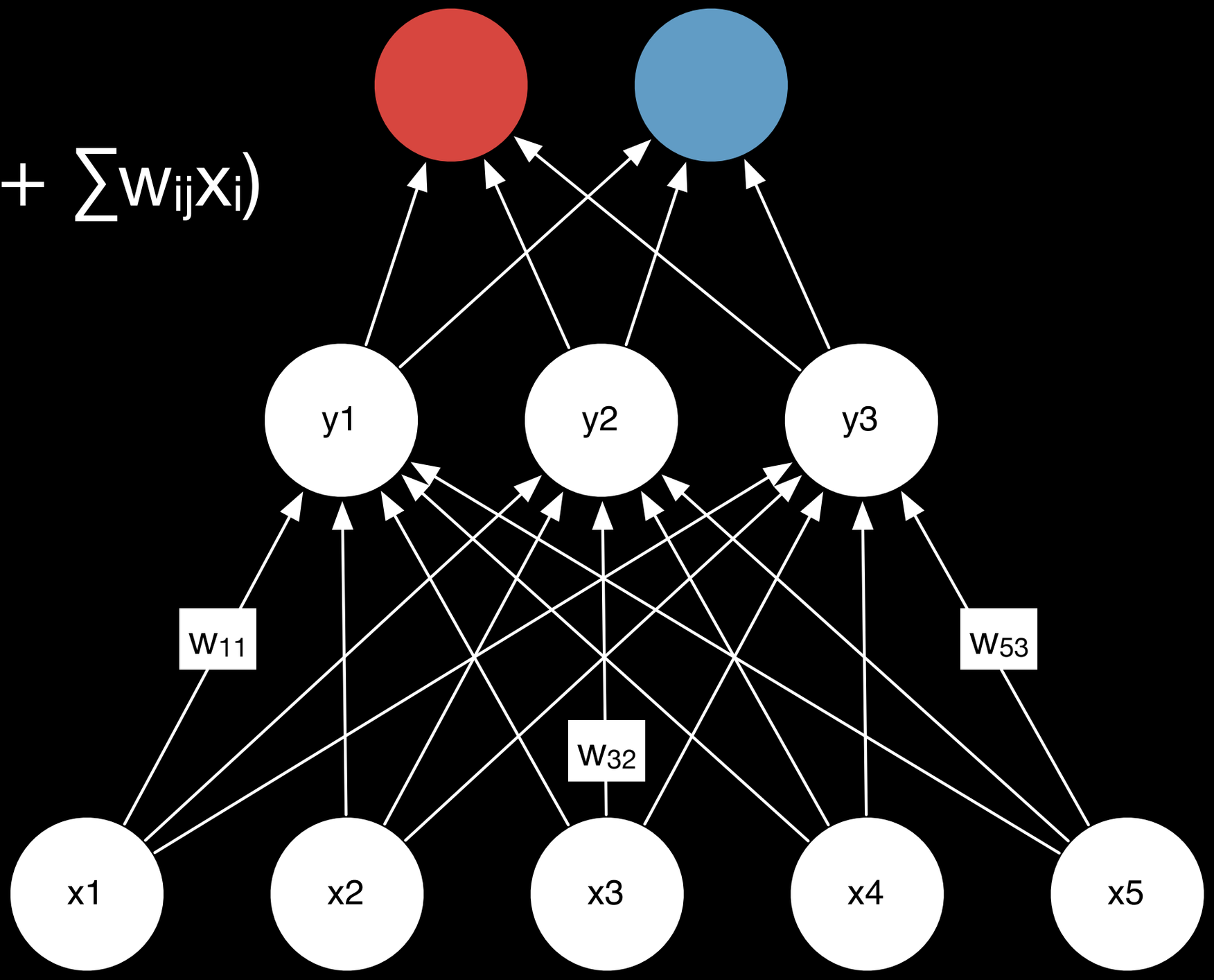

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

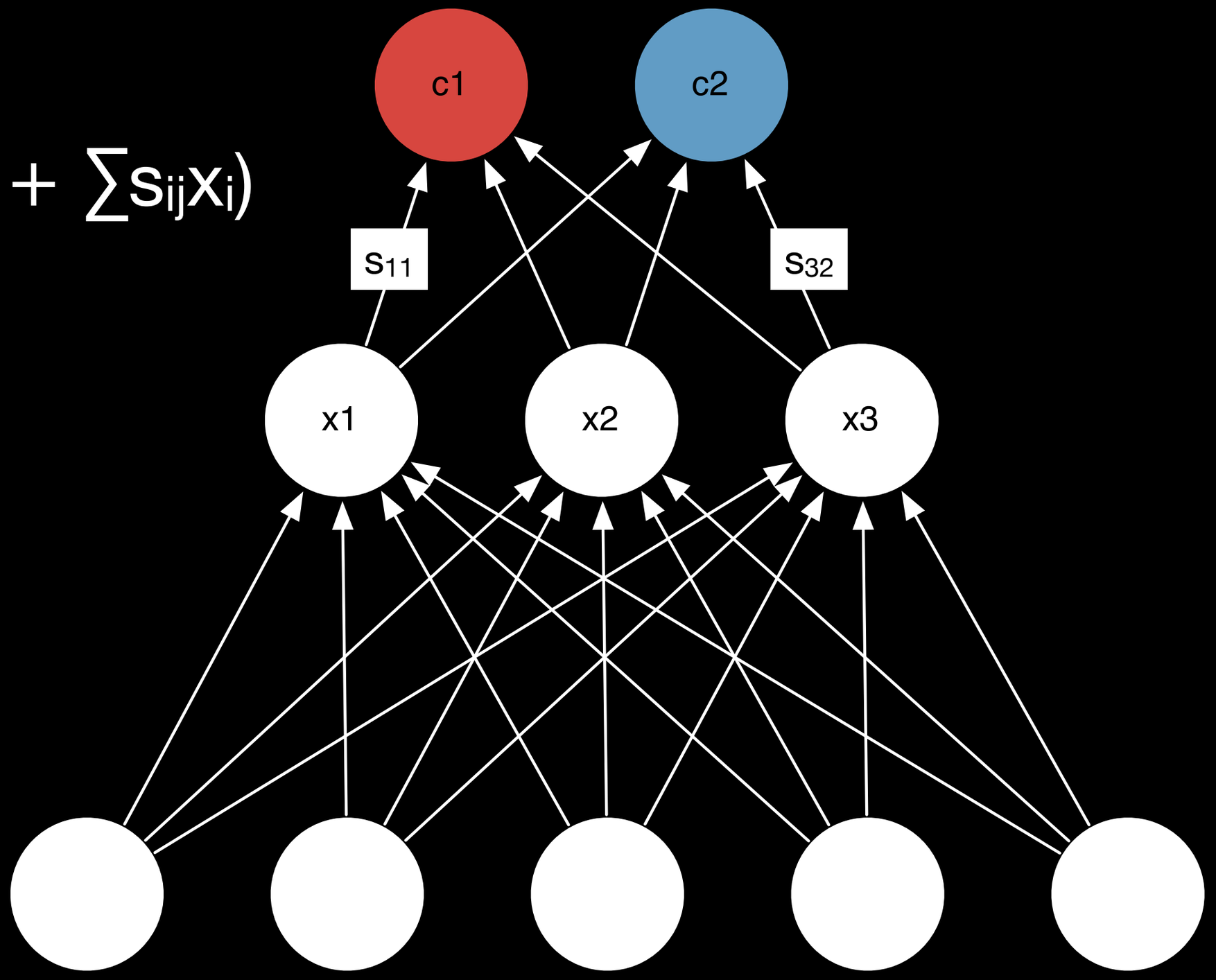
How do neural networks work?

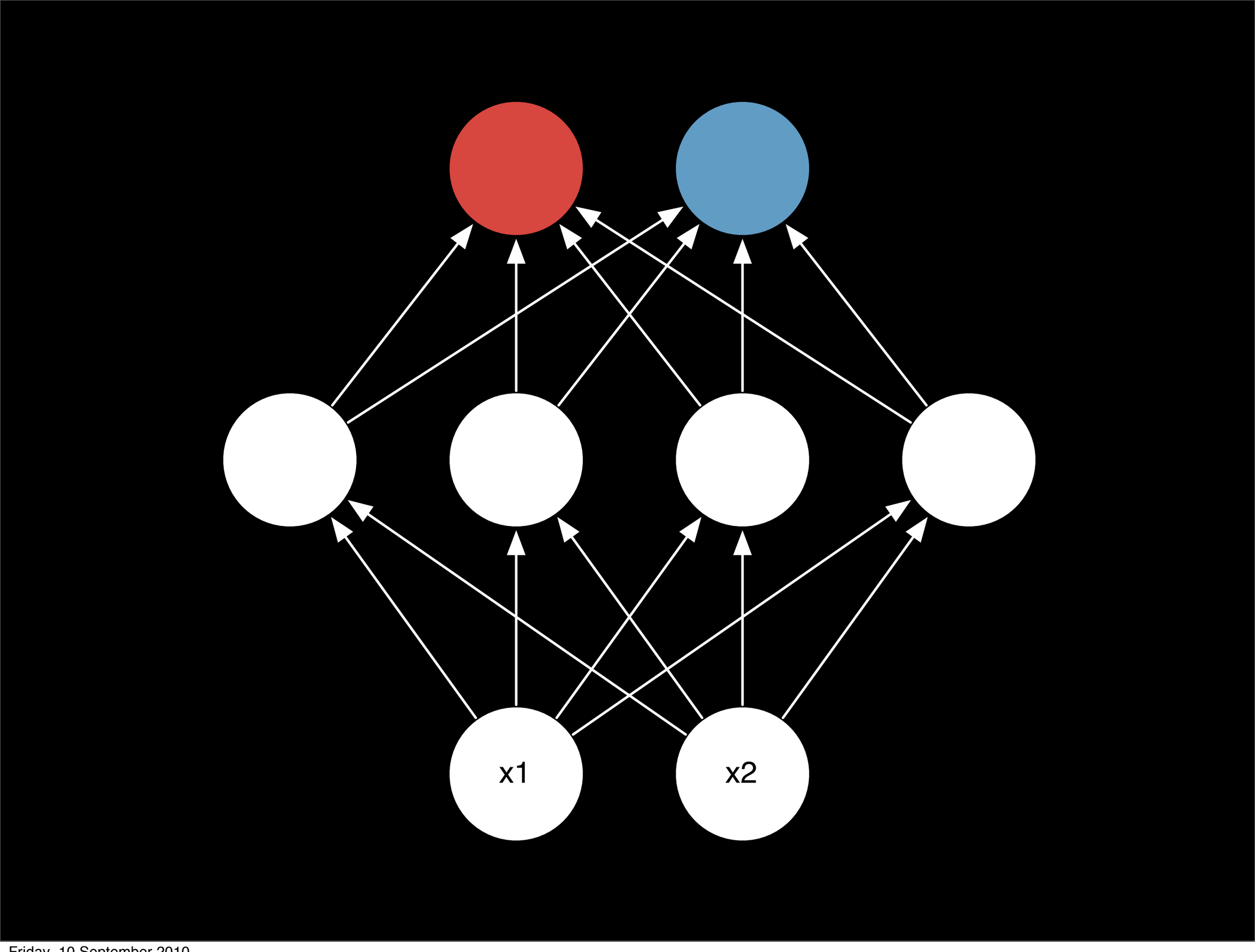

classes
nodes
variables

$$y_j = \text{logit}(\alpha_j + \sum w_{ij} x_i)$$

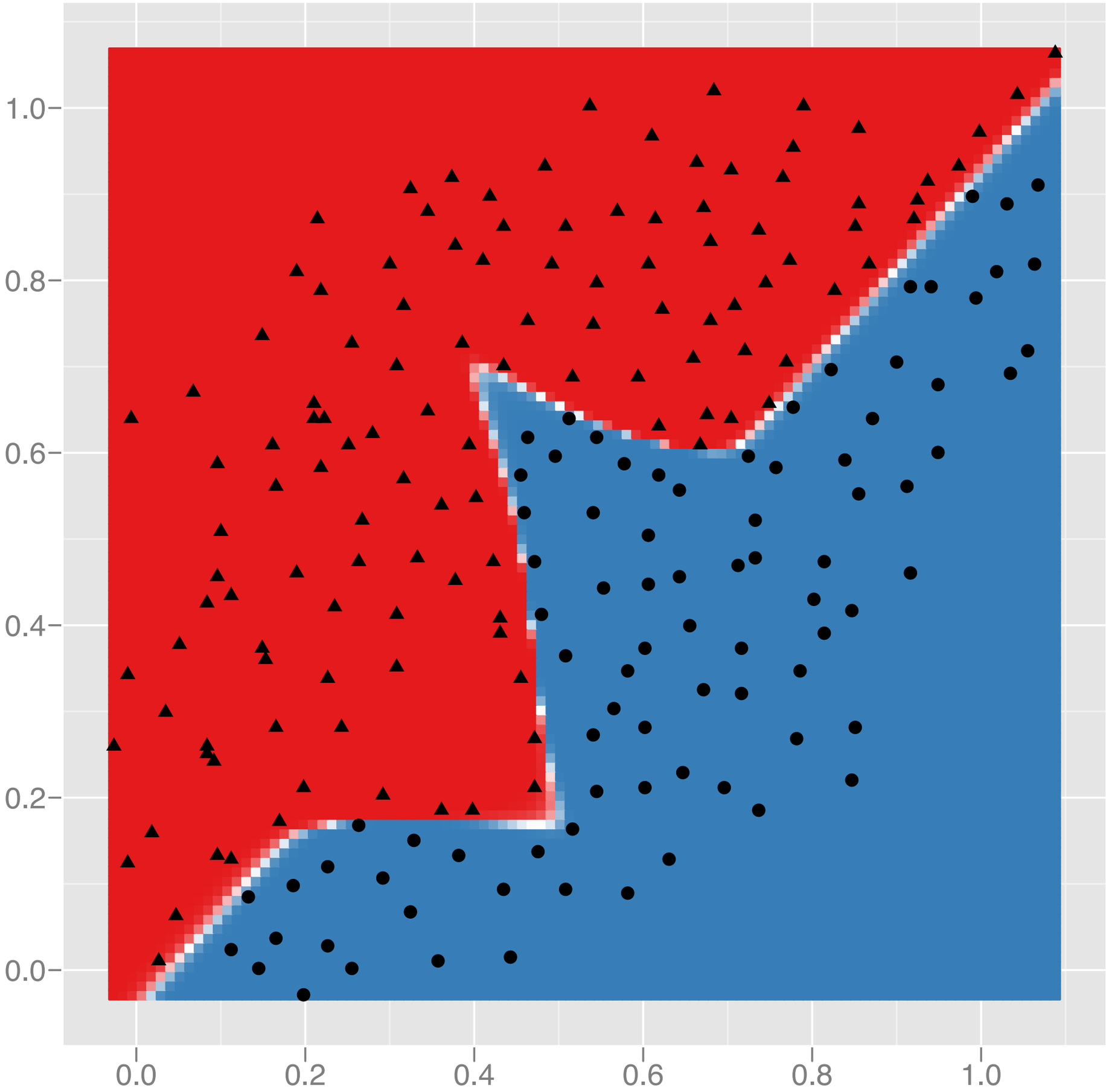


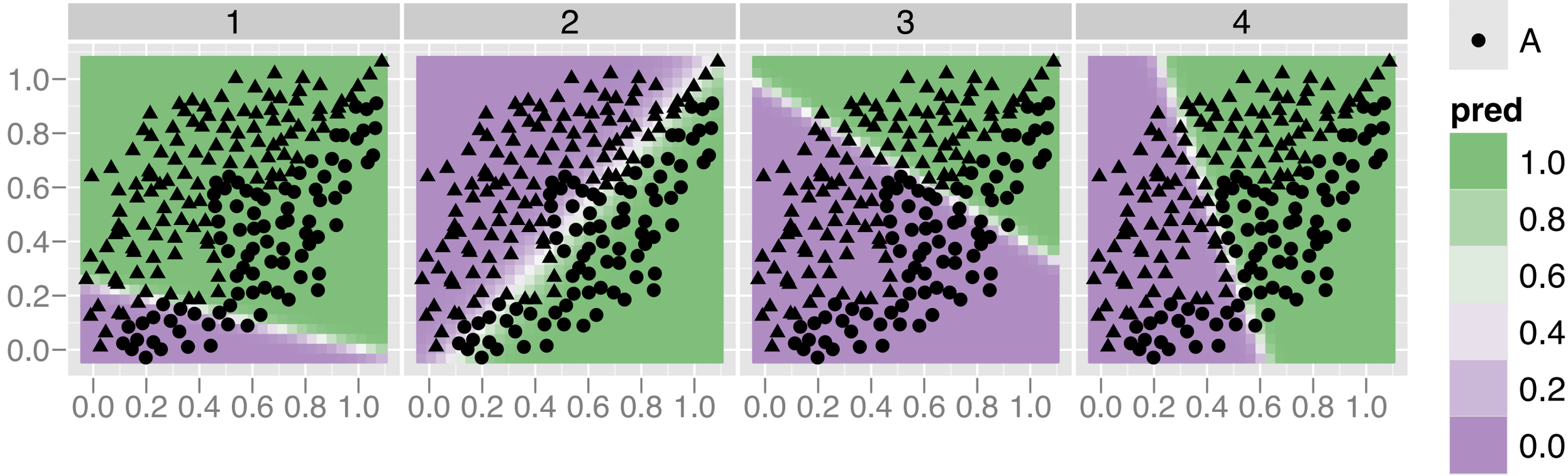
$$c_j = \text{logit}(\alpha_j + \sum s_{ij} x_i)$$

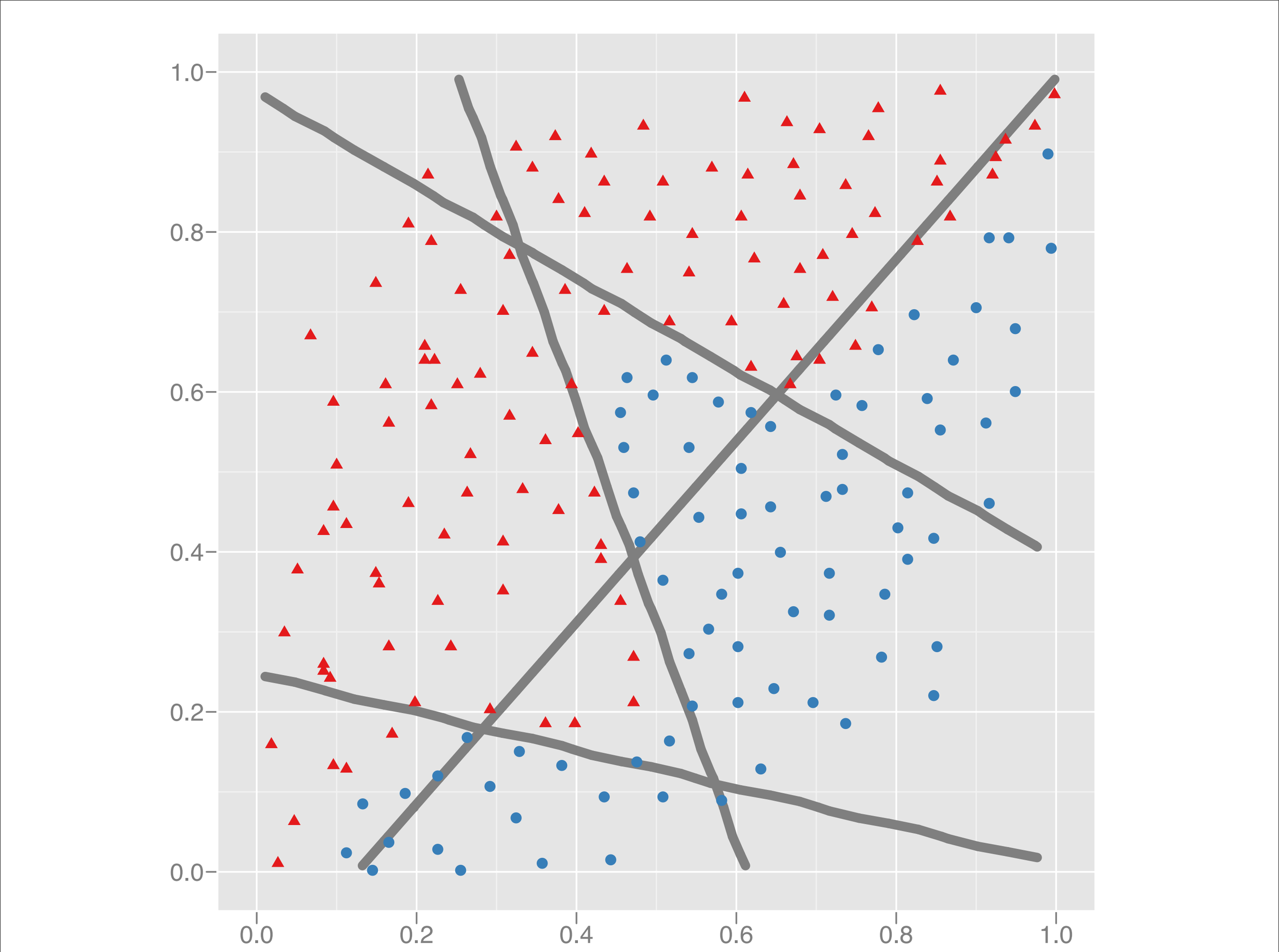


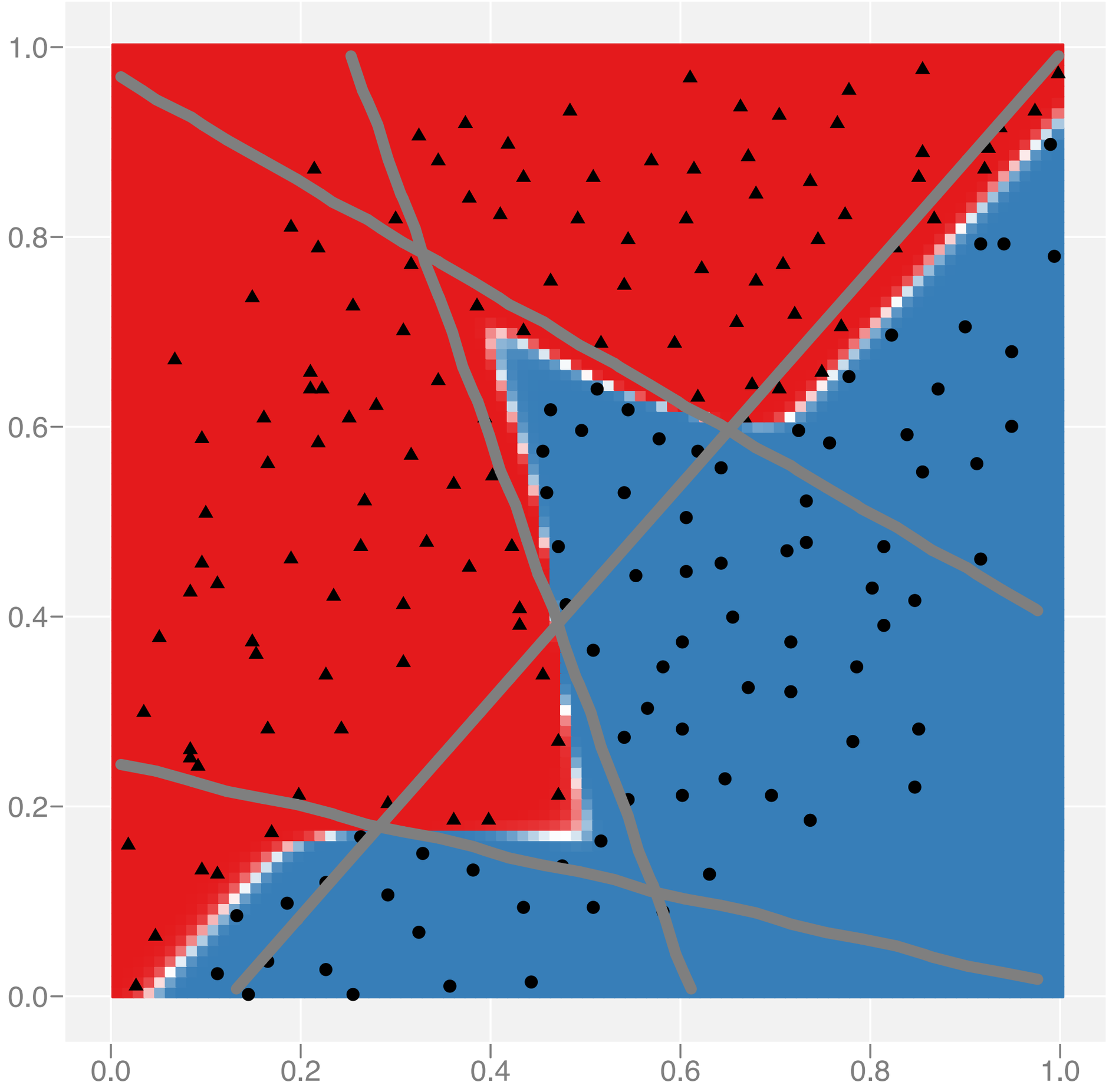

x1
x2

Look at all members of the collection


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1
2
3
4
1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.2 0.4 0.6 0.8 1.0
class
B
A
pred
1.0
0.8
0.6
0.4
0.2
0.0

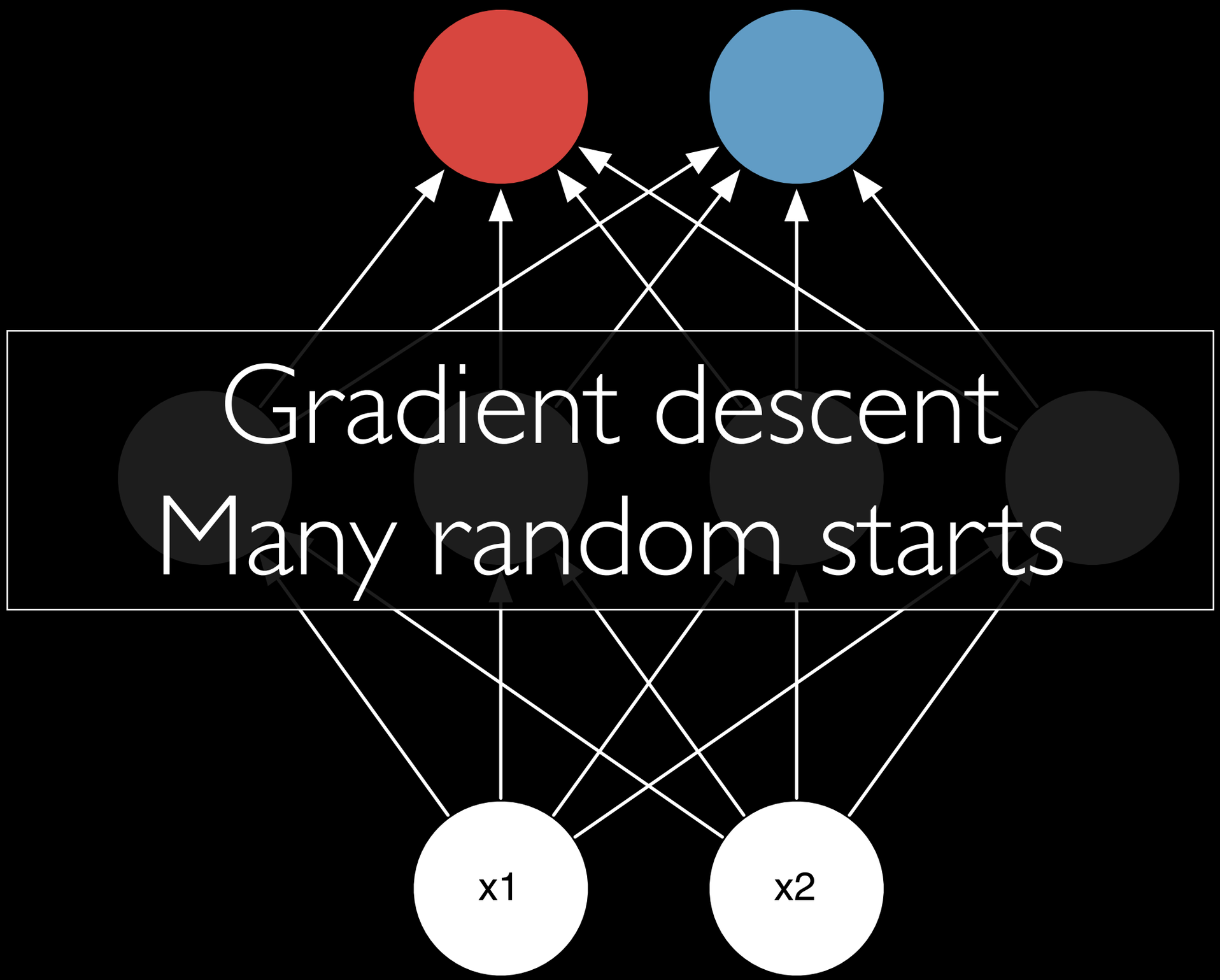

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

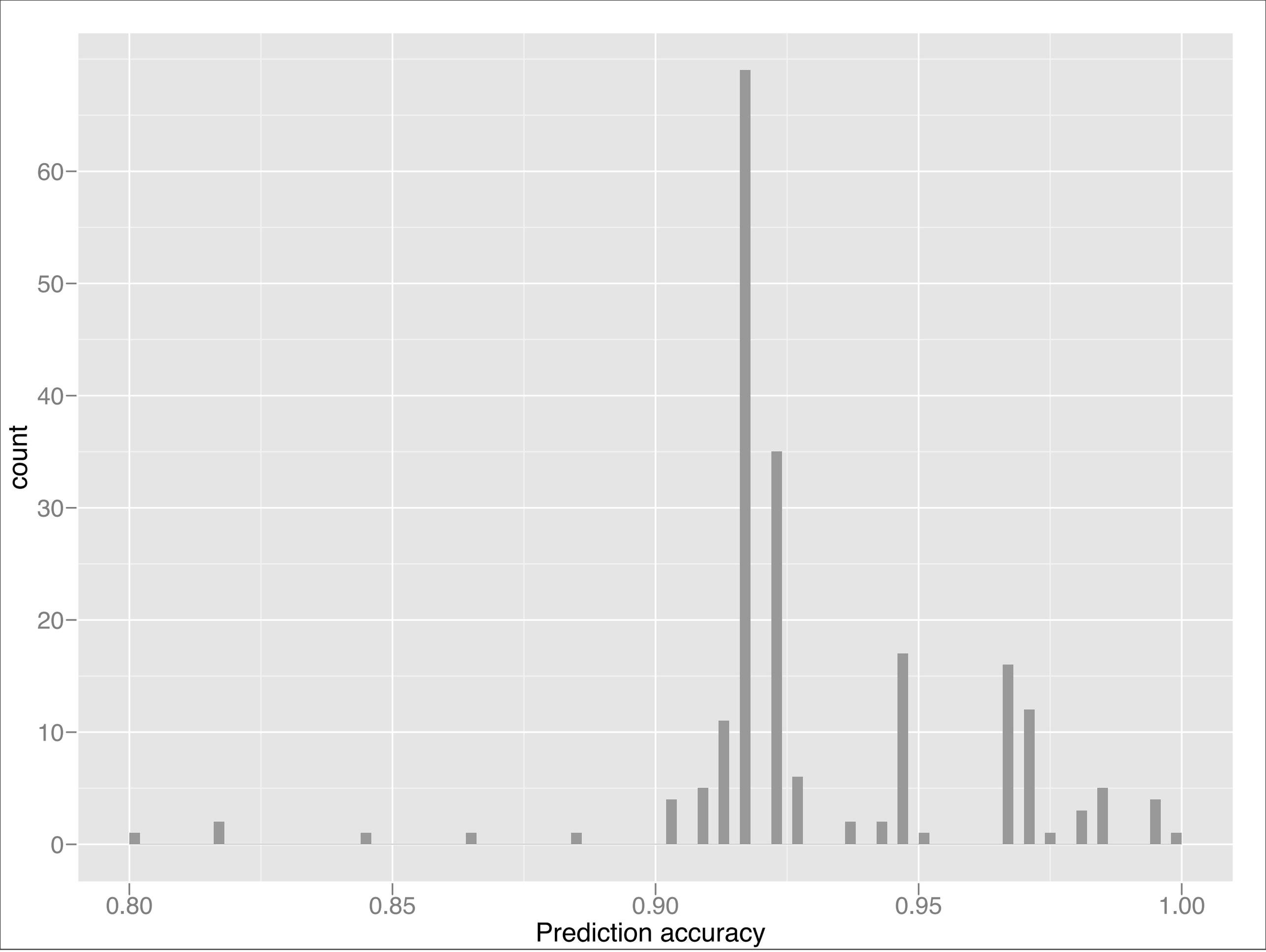

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

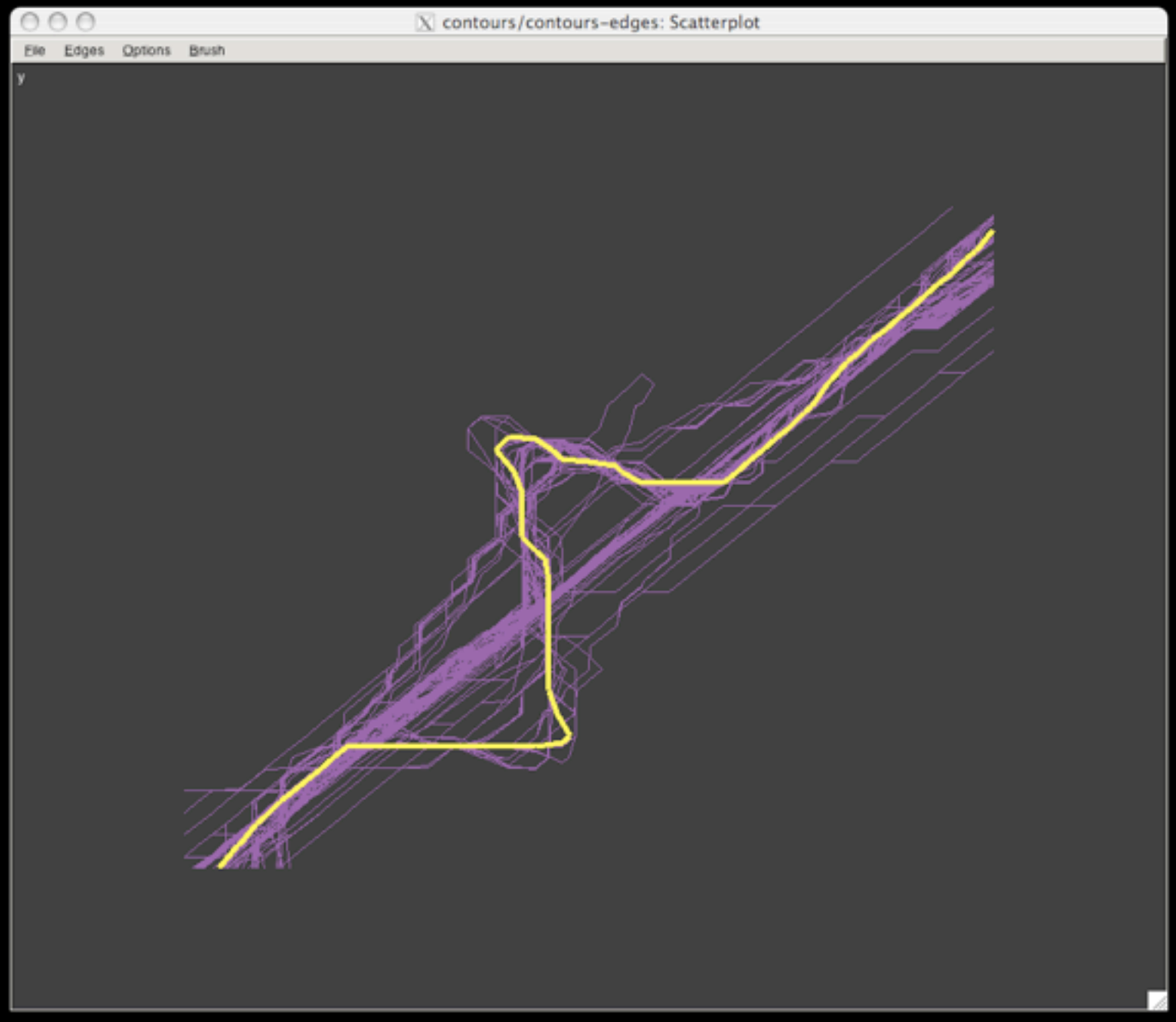
How did I find that model?

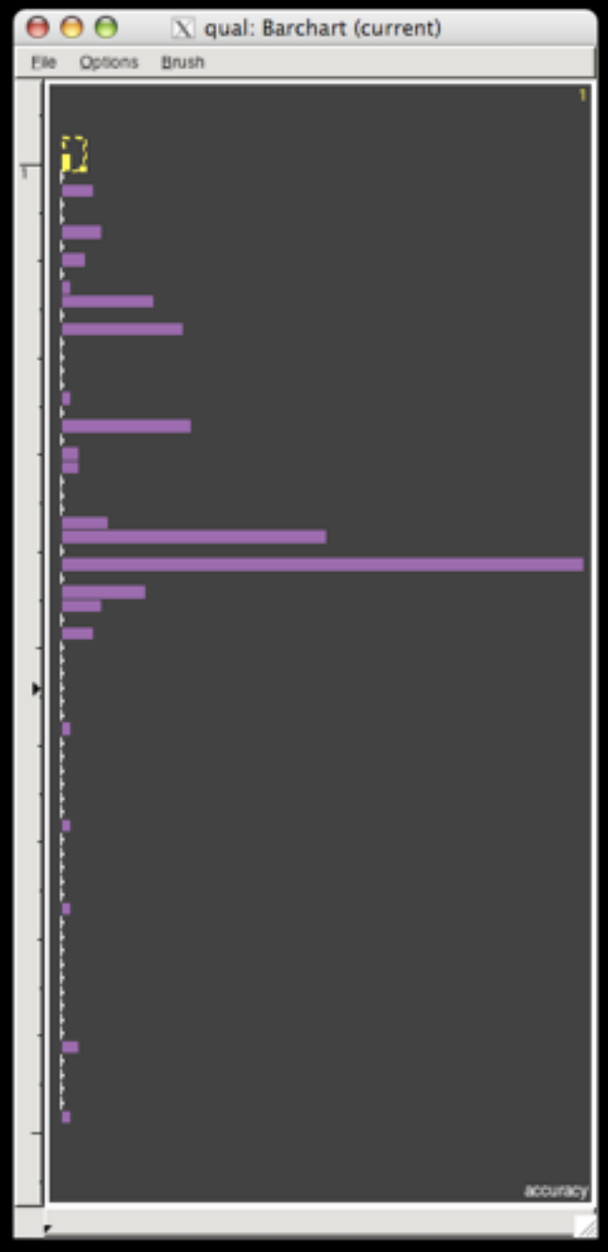
Gradient descent

Many random starts




count
60
50
40
30
20
10
0
0.80
0.85
0.90
0.95
1.00
Prediction accuracy

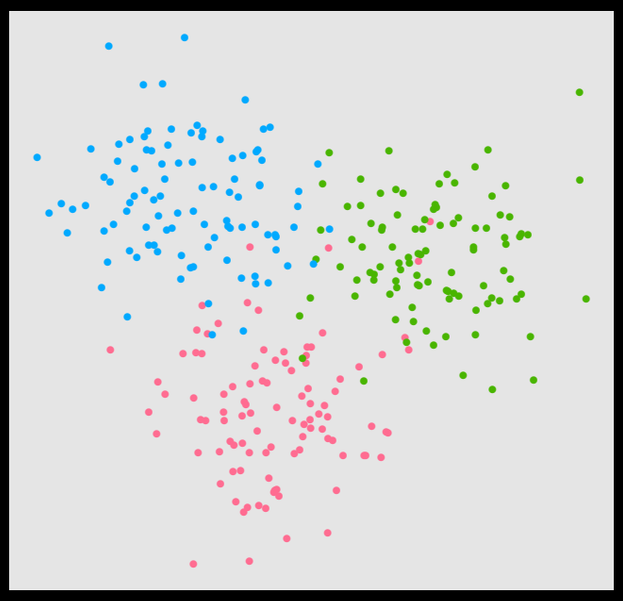

contours/contours-edges: Scatterplot
File Edges Options Brush
y


qual: Barchart (current)
File Options Brush
1
accuracy

Classification algorithm

$f: R^p \rightarrow \{1, 2, ..., k\}$

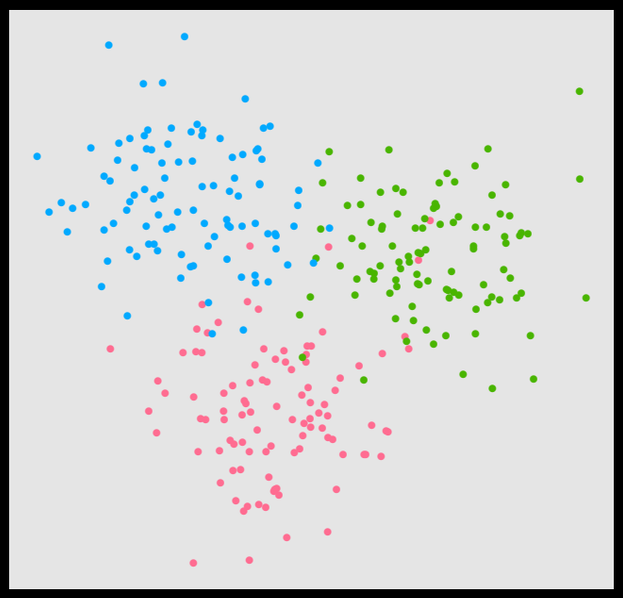
Input



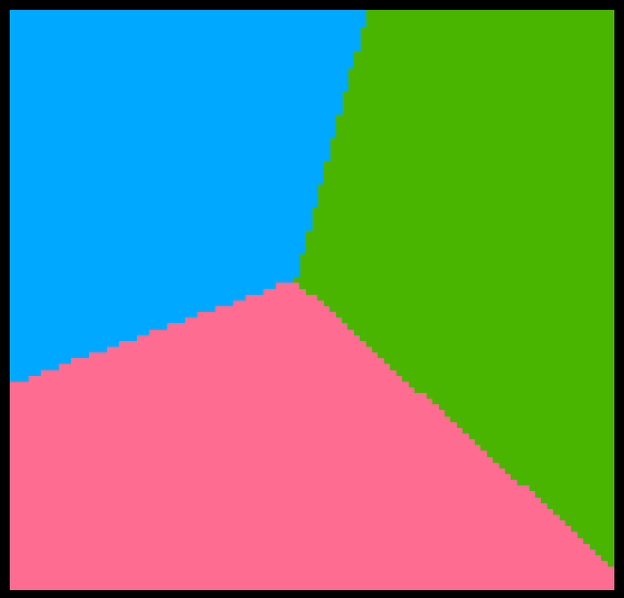
Classification algorithm

$f: R^p \rightarrow \{1, 2, ..., k\}$

Input

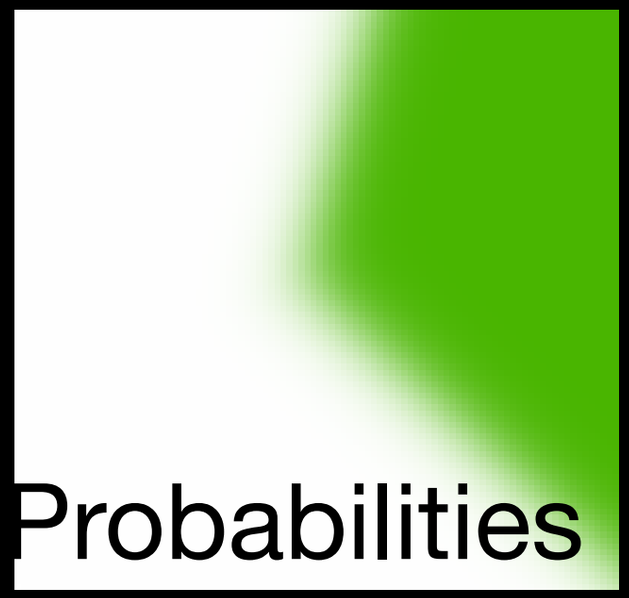


Prediction



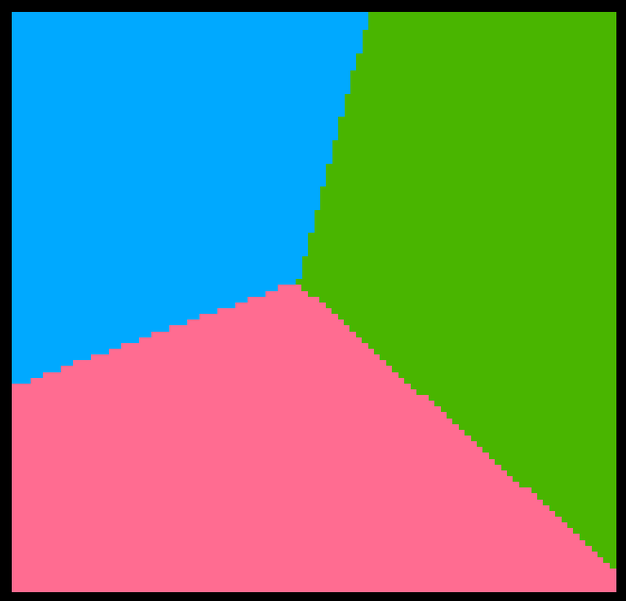
Classification algorithm

$f: R^p \rightarrow \{1, 2, ..., k\}$

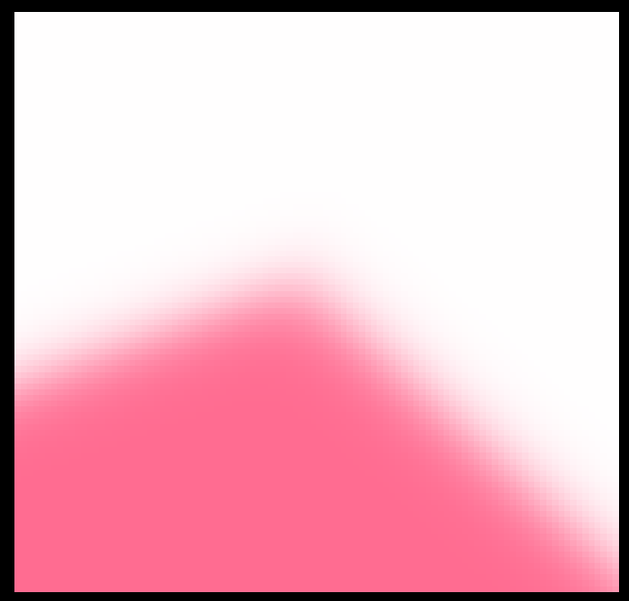


Input

Prediction



Probabilities

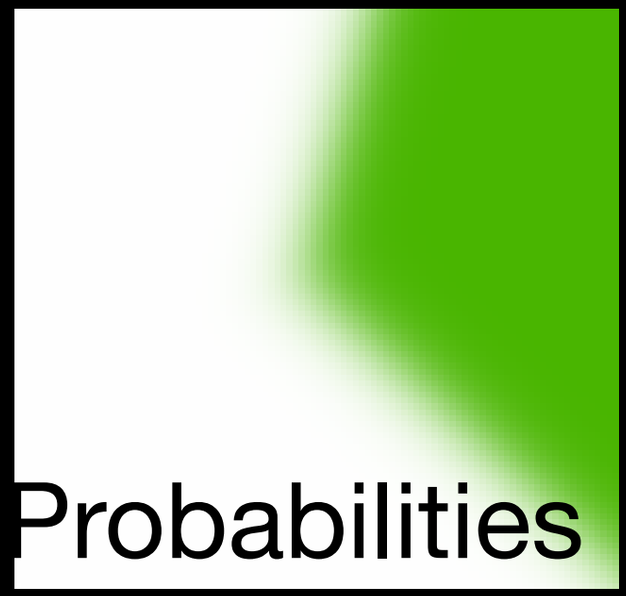


Most also provide class membership probabilities

$f: R^p \rightarrow [0,1]^k$

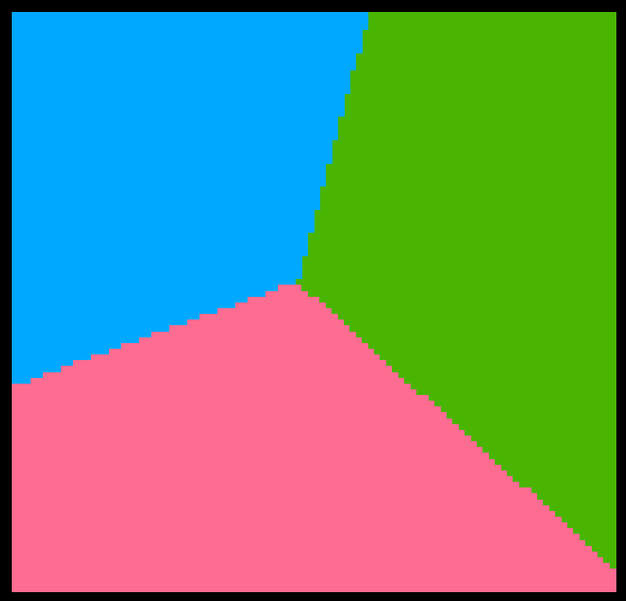
Classification algorithm

$f: R^p \rightarrow \{1, 2, ..., k\}$



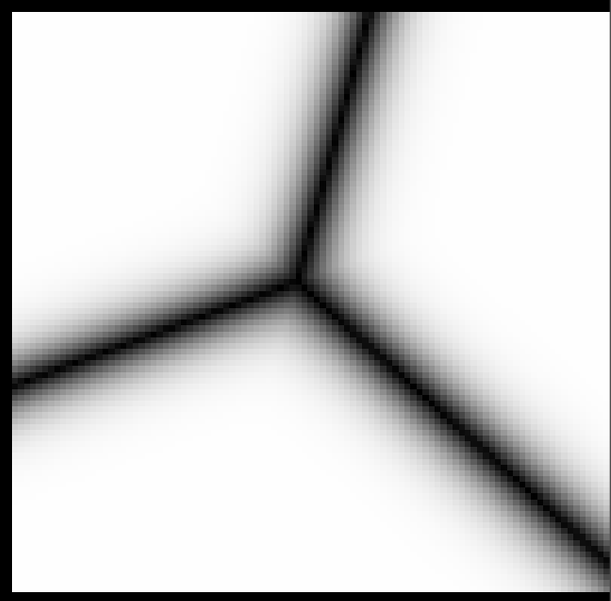
Input

Prediction



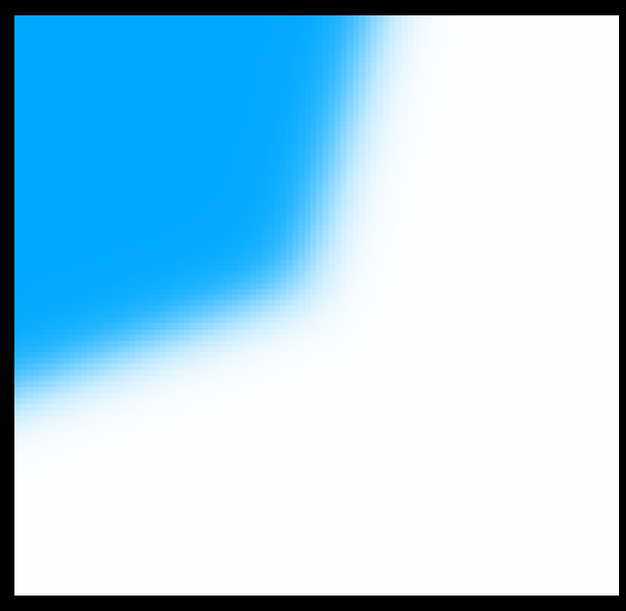
Probabilities

Advantage



Most also provide class membership probabilities

$f: R^p \rightarrow [0,1]^k$

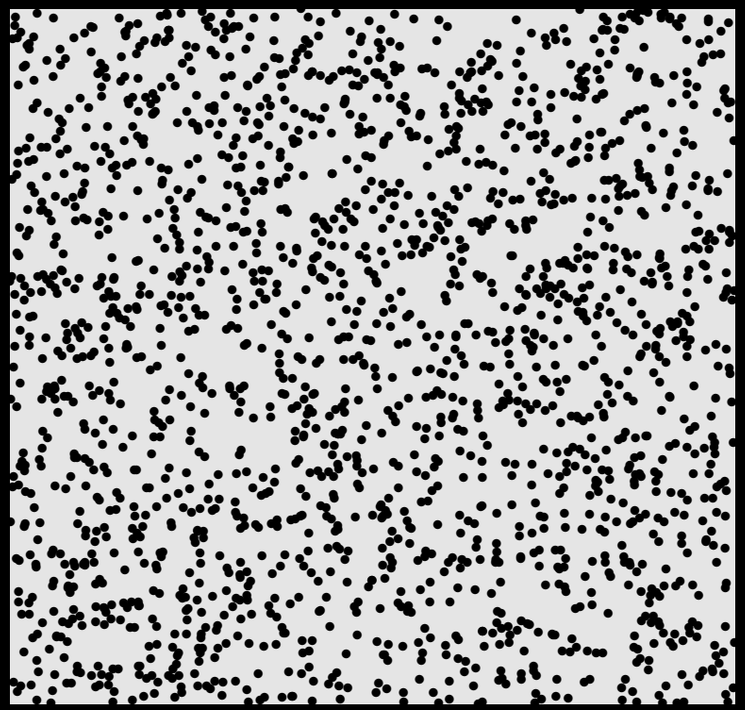


P(best) - P(second best)

How to find the boundaries?

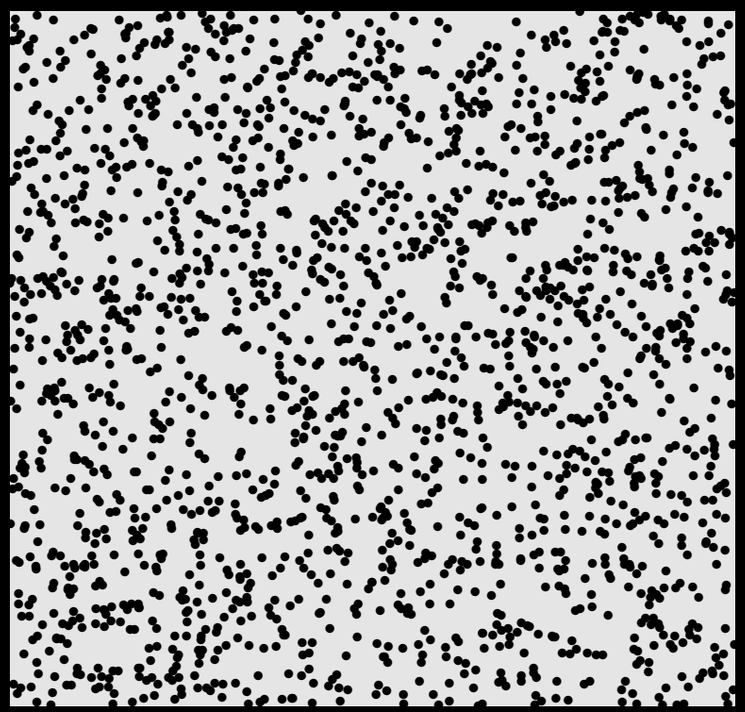
How to find the boundaries?

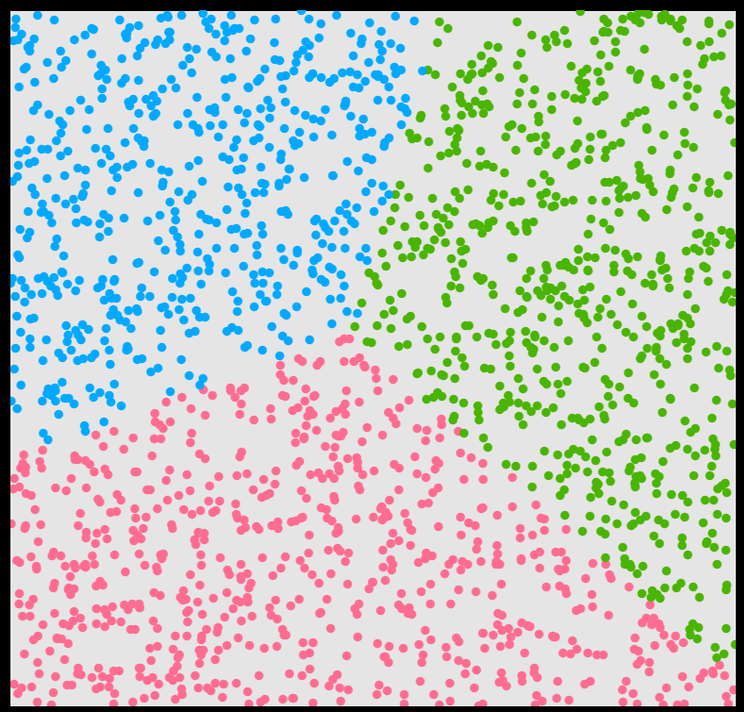
Random sample



How to find the boundaries?

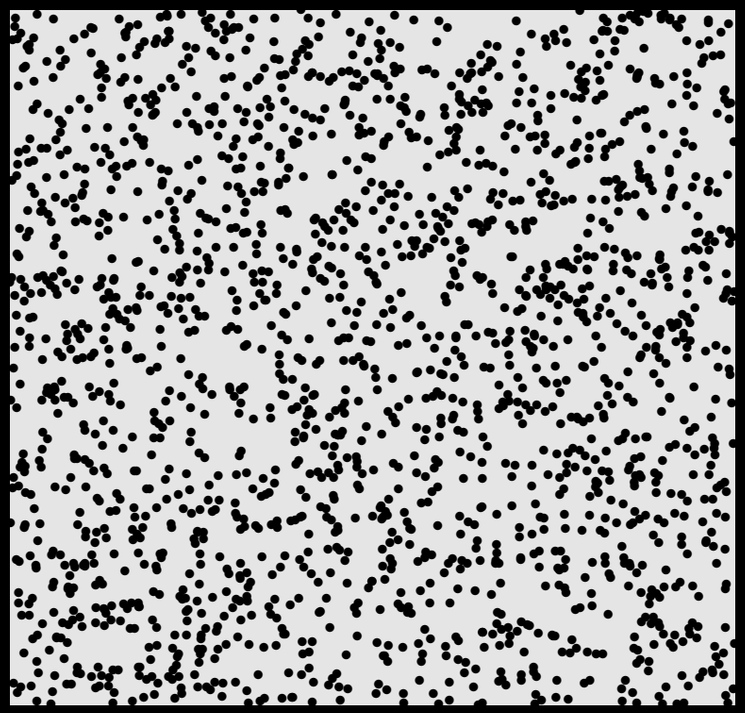
Random sample Classify



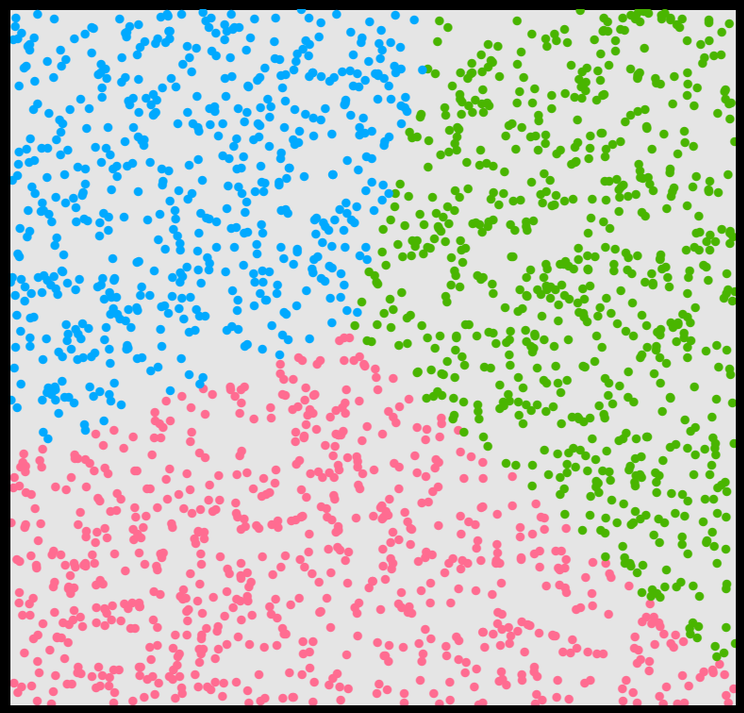


How to find the boundaries?

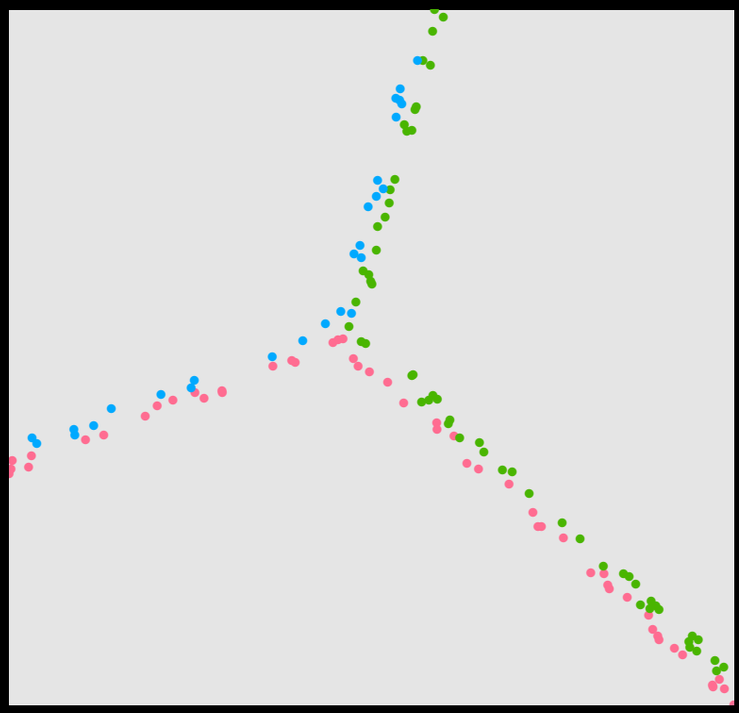
Random sample



Classify

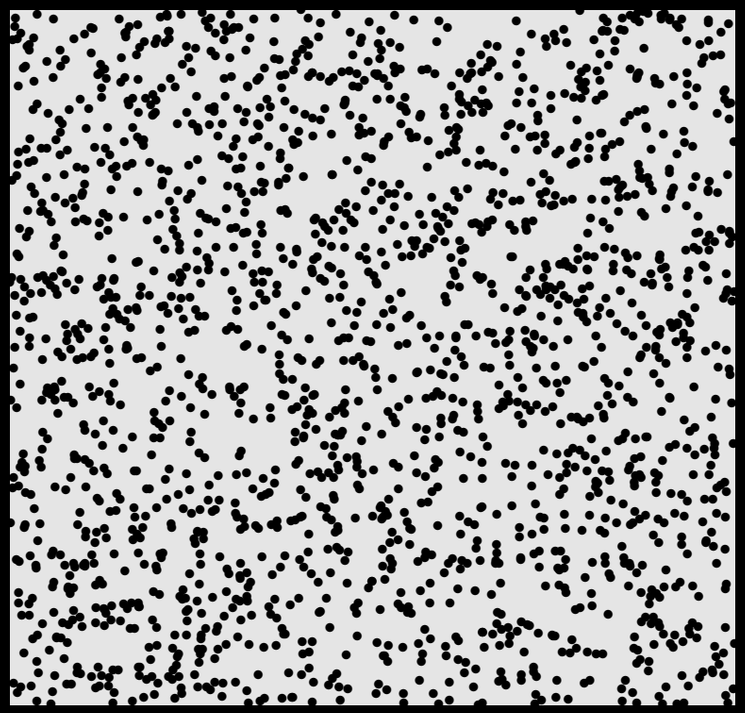


Low advantage

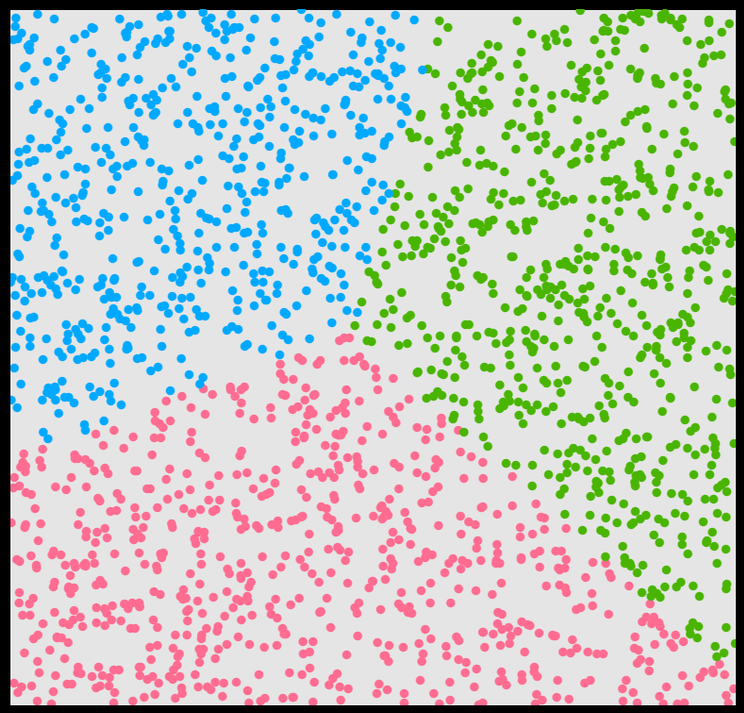


How to find the boundaries?

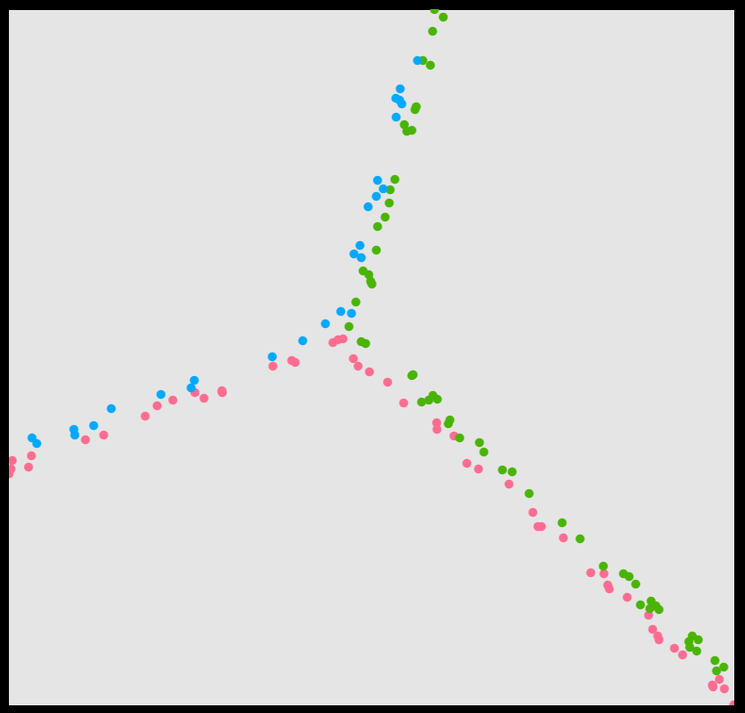
Random sample



Classify



Low advantage



Crude method, but works for all classification algorithms and for moderate dimensionality

Ensembles of linear models

Display the model in the data space

Look at many members of a collection

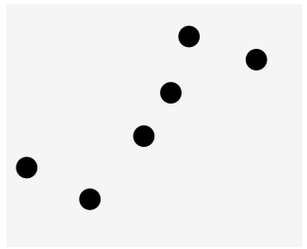
Explore the process of fitting, not just the end result

Data

- Fertility in French-speaking Swiss provinces in the late 1800's
- Predict fertility based on:
  - proportion of agricultural workers
  - average performance on an army examination
  - amount of higher education
  - proportion of Catholics
  - infant mortality

Model

- Linear modes with all combinations of covariates (2^p models)
- What can looking at all models tell us that looking at just a few can't?



Observation

Obs ID

Original data

Model-observation summaries

1

many

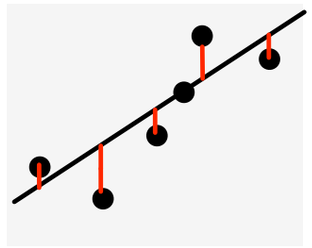
Model-Observation

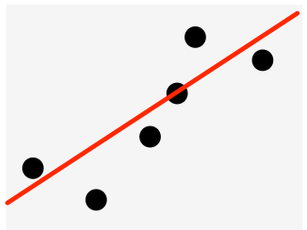
Obs ID

Model ID

Diagnostics

Fit quality





Model

Model ID

Model fit statistics

1

many

1

many

Model-Estimate

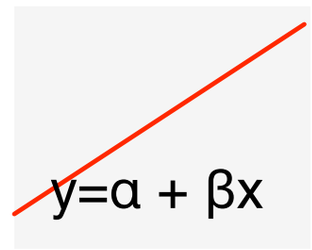
Model ID

Estimate ID

Raw

Standardised

Uncertainty



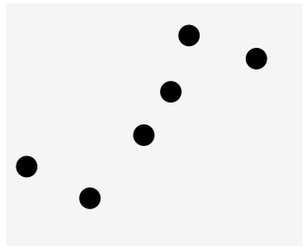
f(**α**, **β**)

Estimate

Estimate ID

n

model-estimate summaries



Observation

Obs ID

Original data

Model-observation summaries

1

many

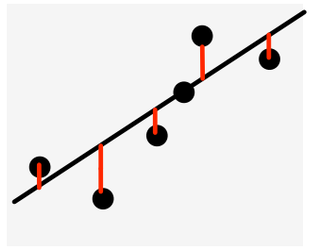
Model-Observation

Obs ID

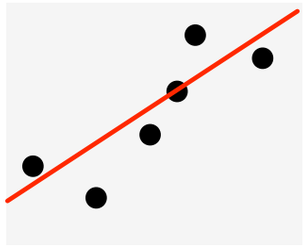
Model ID

Diagnostics

Fit quality



many



Model

Model ID

Model fit statistics

1

1

many

Model-Estimate

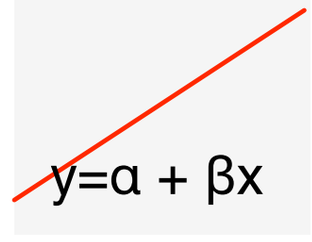
Model ID

Estimate ID

Raw

Standardised

Uncertainty



f(**α**, **β**)

Estimate

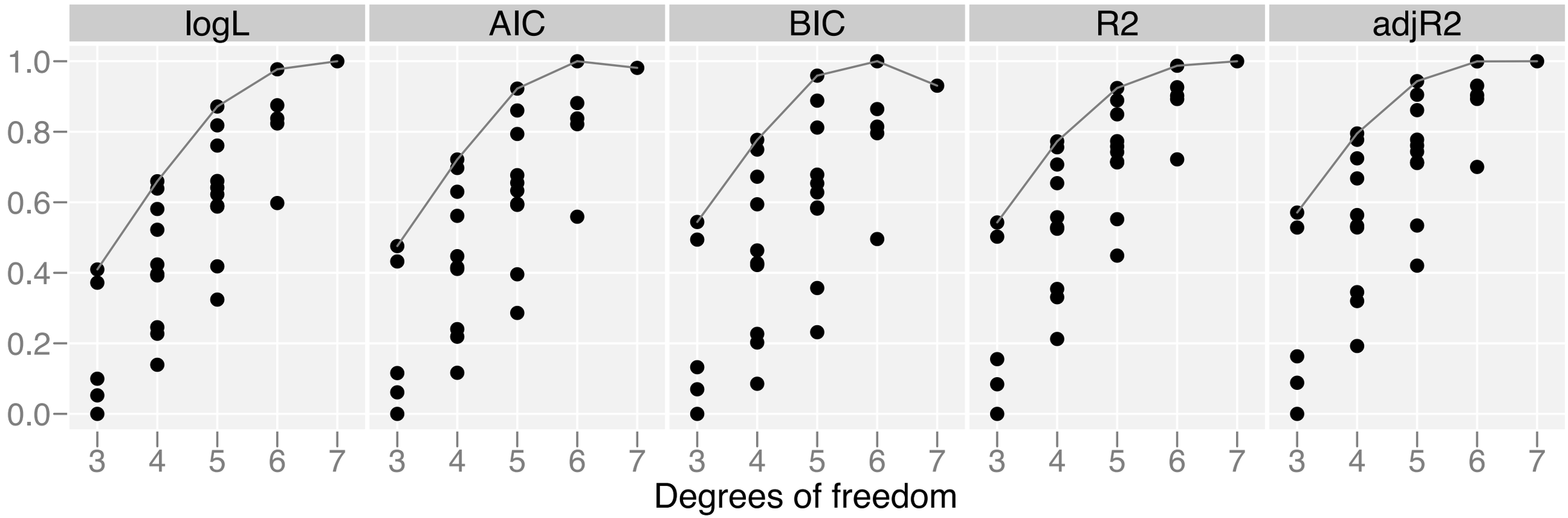
Estimate ID

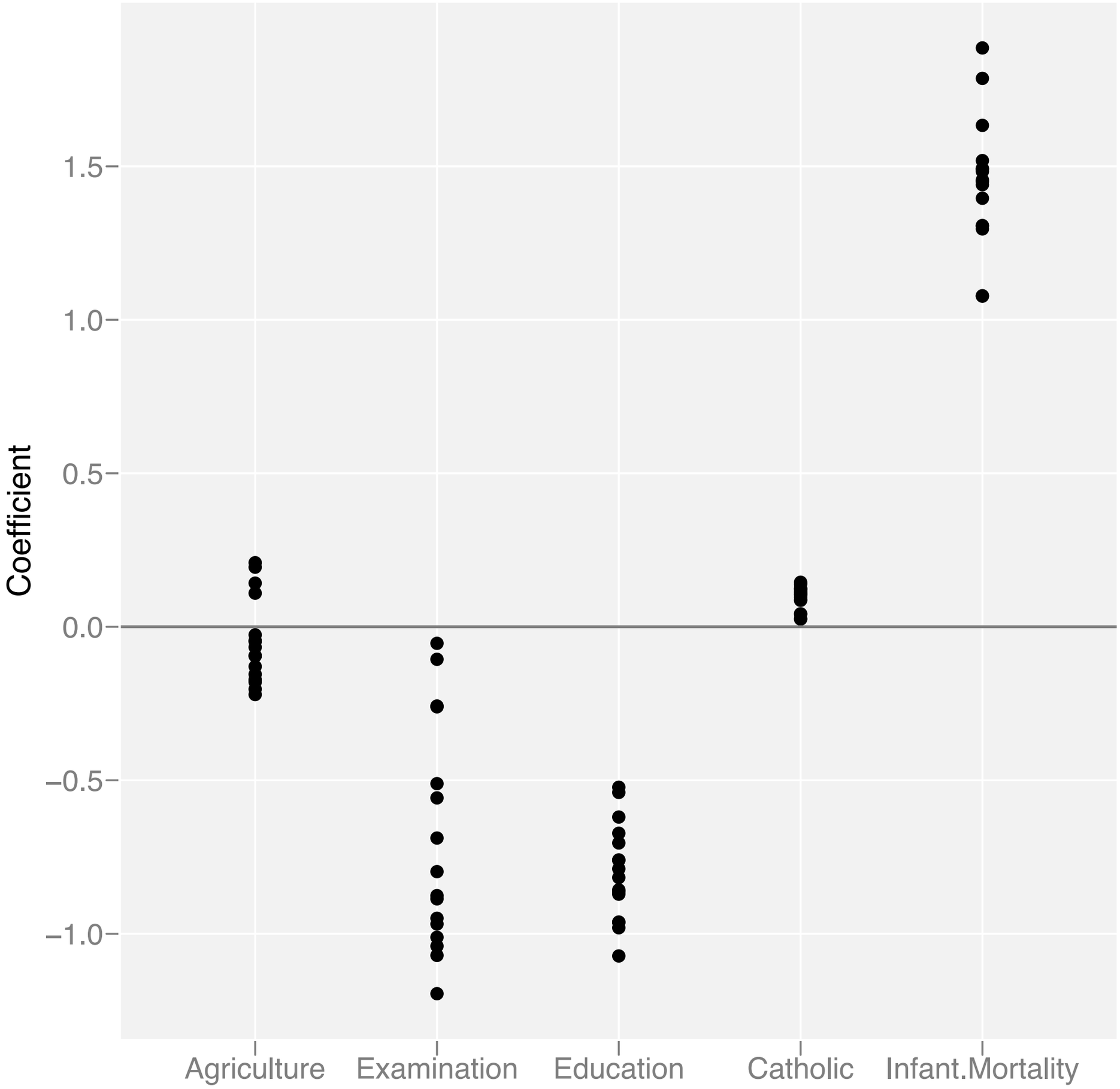
n

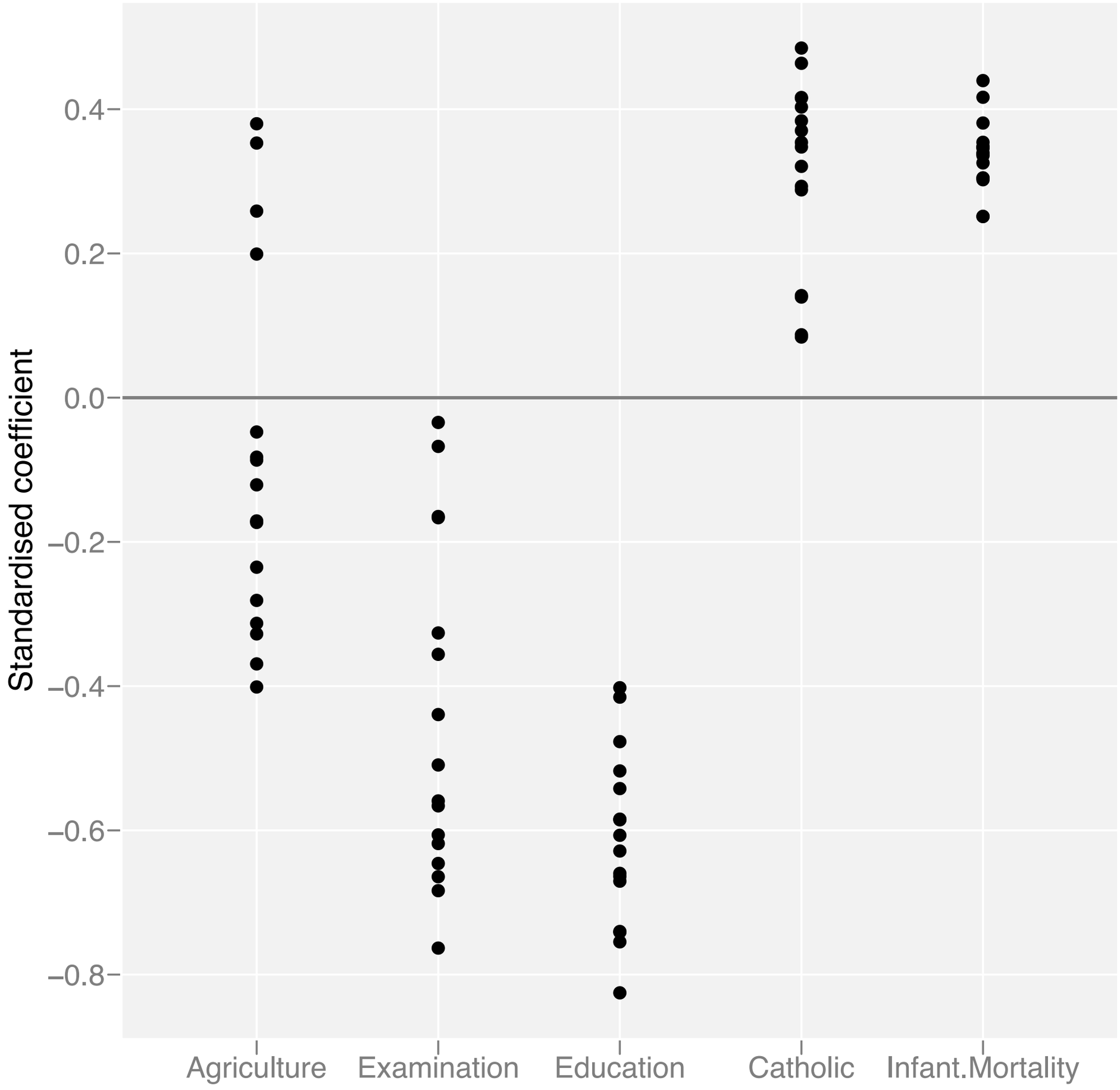
model-estimate summaries

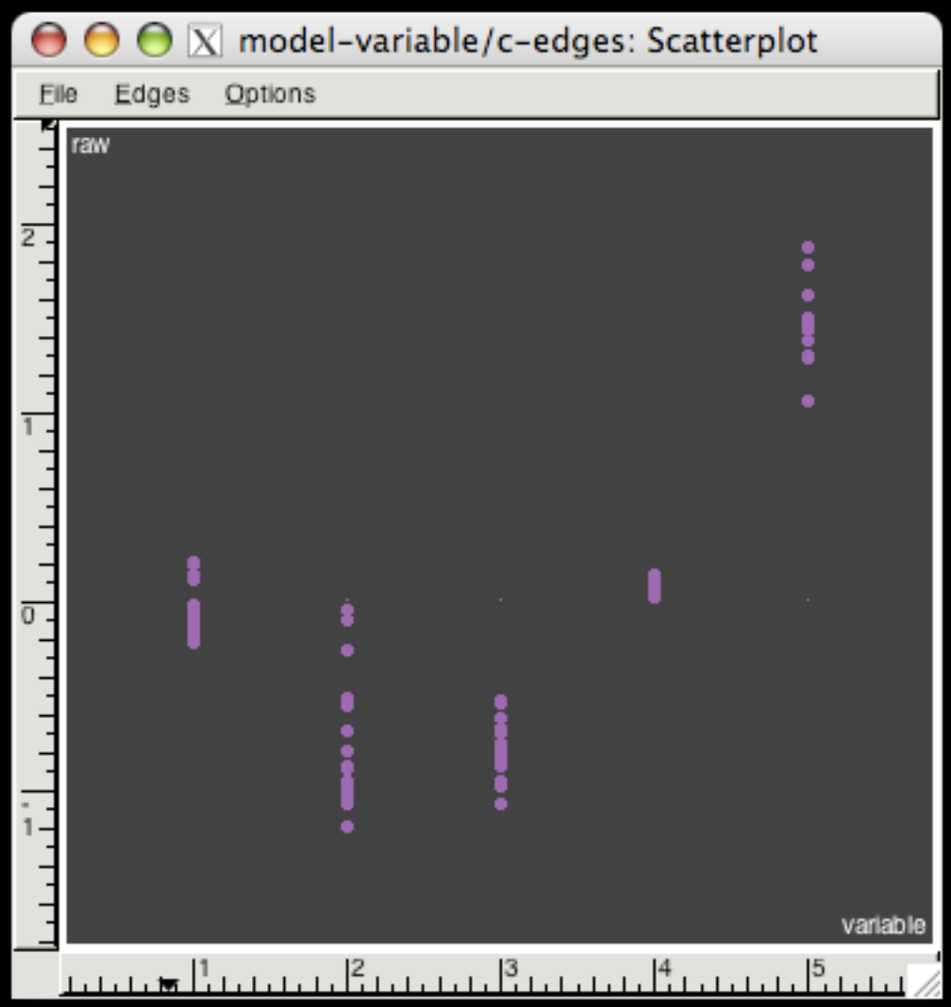
1

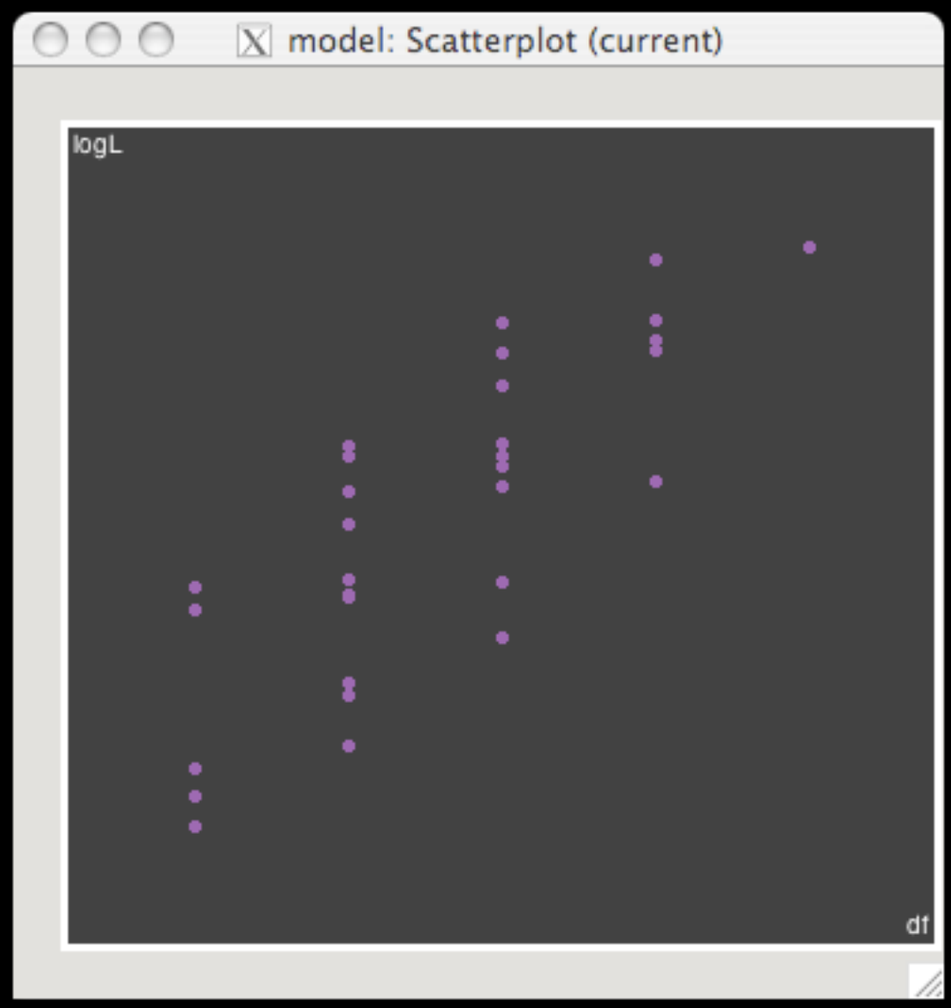
many


logL
AIC
BIC
R2
adjR2
1.0
0.8
0.6
0.4
0.2
0.0
3
4
5
6
7
Degrees of freedom


Coefficient
1.5
1.0
0.5
0.0
−0.5
−1.0
Agriculture
Examination
Education
Catholic
Infant.Mortality


Standardised coefficient
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
Agriculture
Examination
Education
Catholic
Infant.Mortality


model-variable/c-edges: Scatterplot
File
Edges
Options
raw
variable
2
1
0
1
1
2
3
4
5


model: Scatterplot (current)
logL
df

Conclusions

Other methods

- MANOVA
- Self-organising maps (clusterfly)
- Hierarchical clustering (clusterfly)
- Classification methods (classifly)
- Projection pursuit (tourr)

The future

- Better iteration between modelling and visualisation
- Foundations to make interactive graphics easy to produce in R