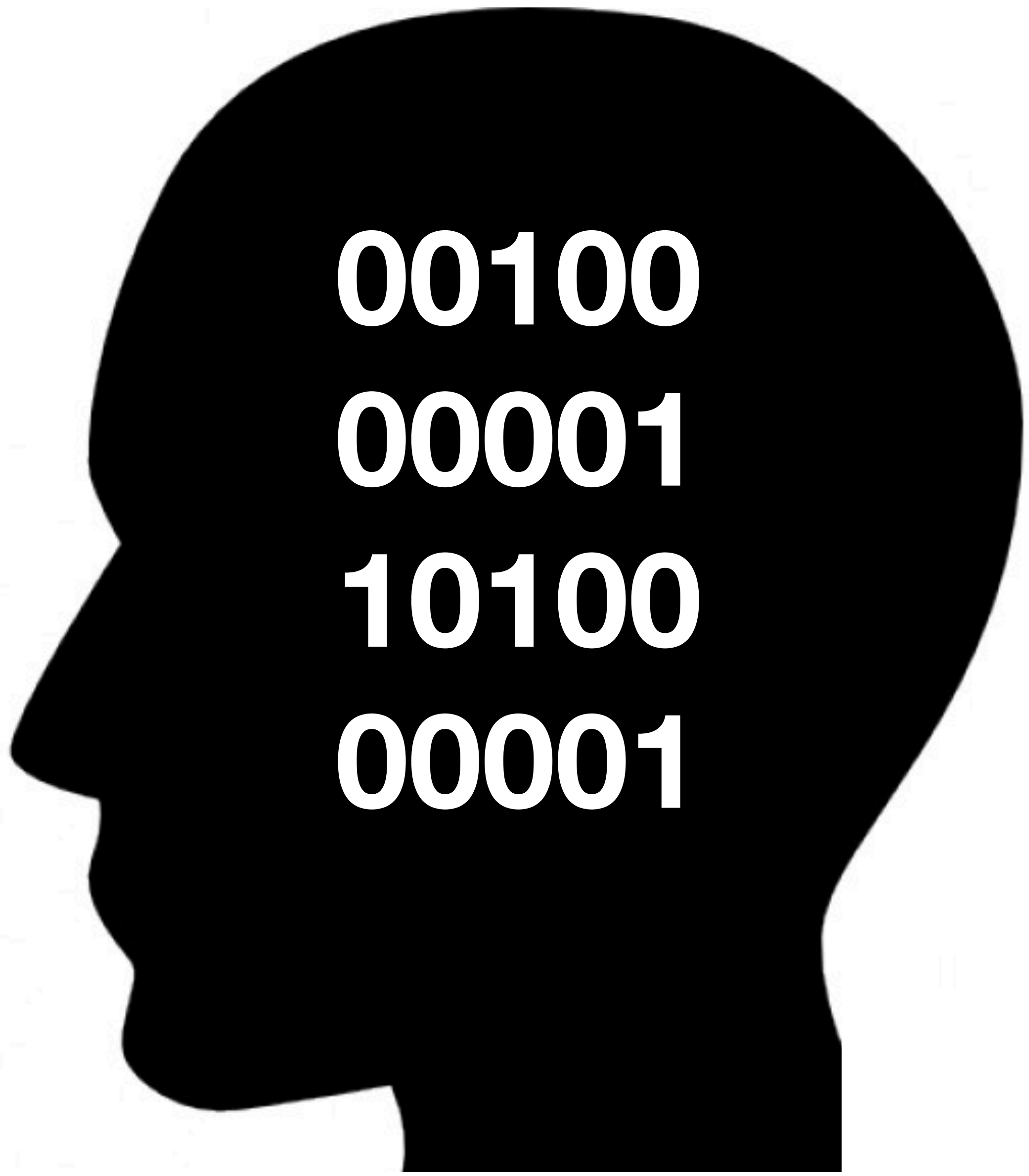
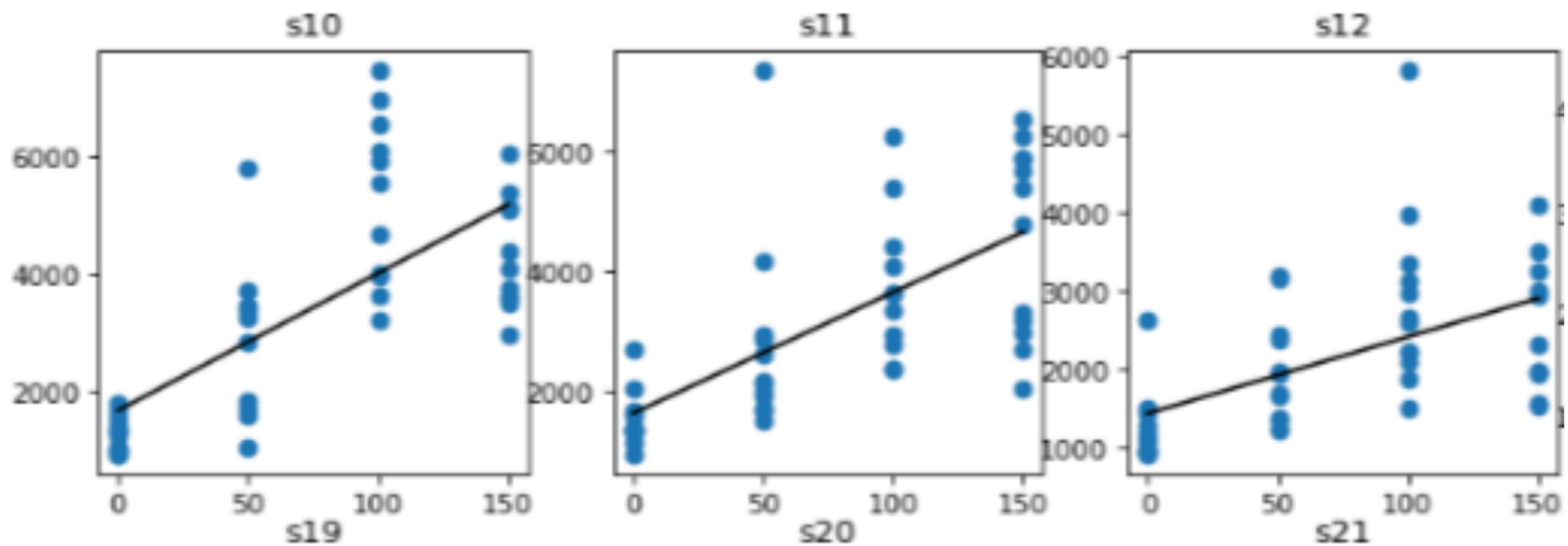


COG260: Data, Computation, and The Mind

Categorization

Lab 3: Mental rotation



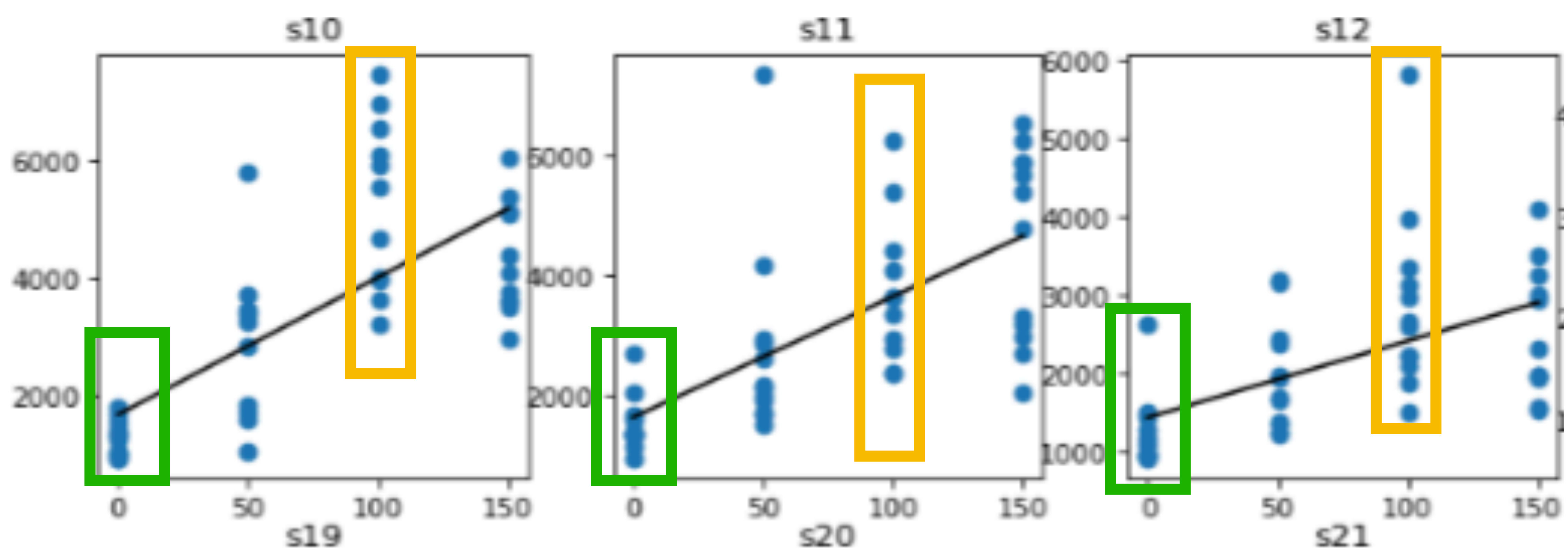
RT (msec)

Angle

Lab 3: Mental rotation

Response variation across different stimuli pairs

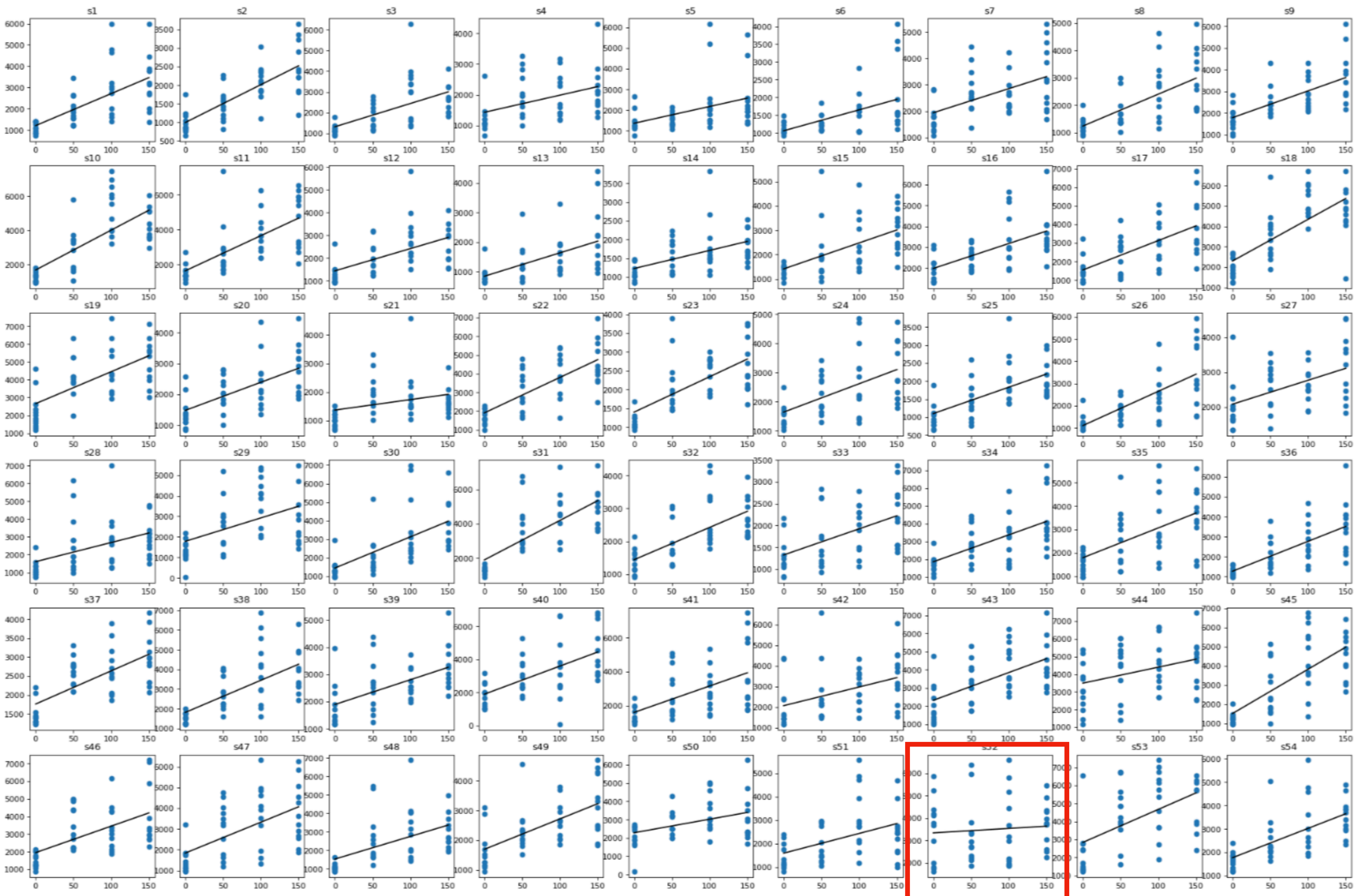
RT (msec)



Angle

Lab 3: Mental rotation

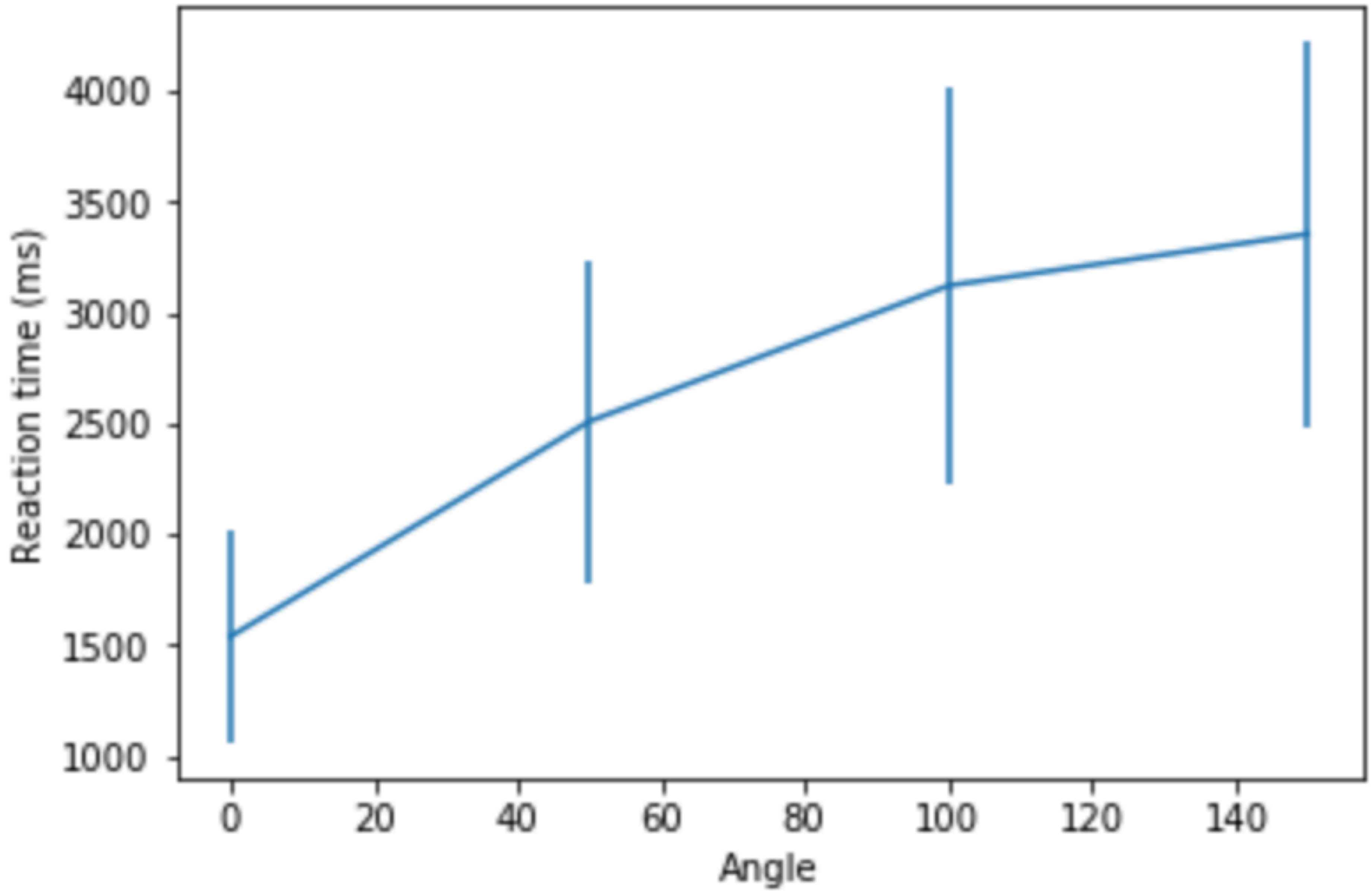
Reaction time (ms)



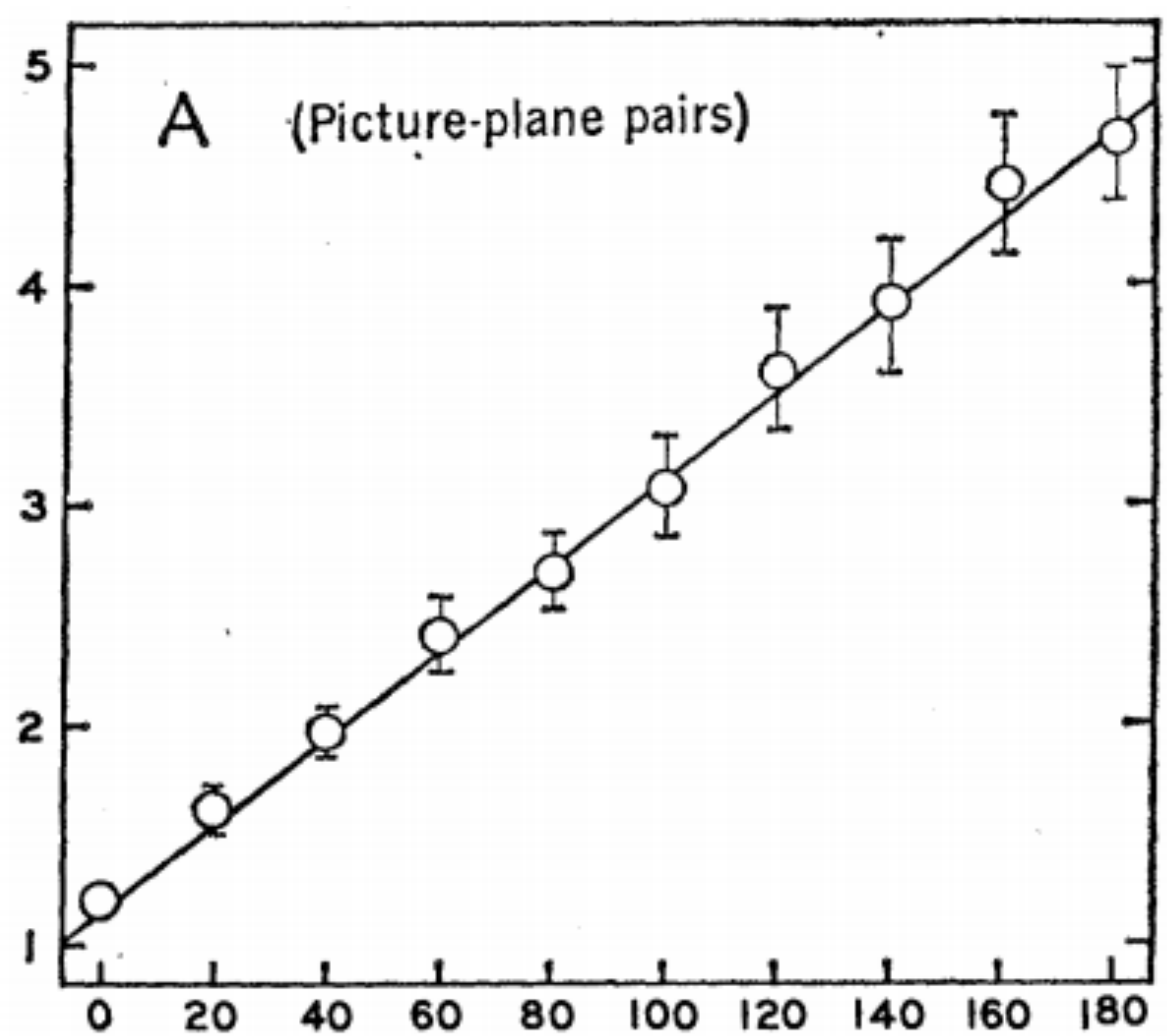
Weak effect (individual variation)

Lab 3: Mental rotation

Group-level statistics support the hypothesis



Comparison to Shepard & Metzler (1971)

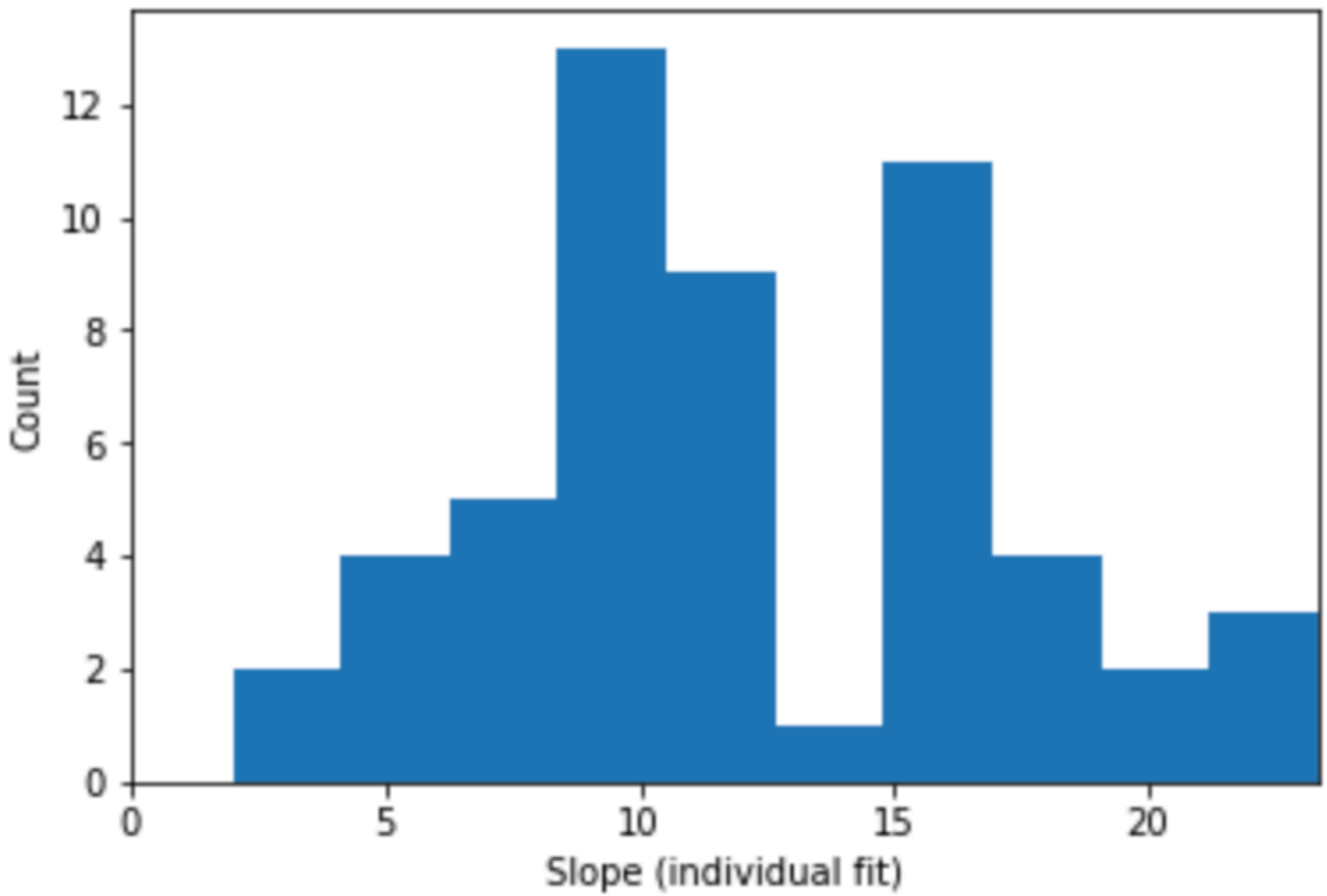


RT

Angle

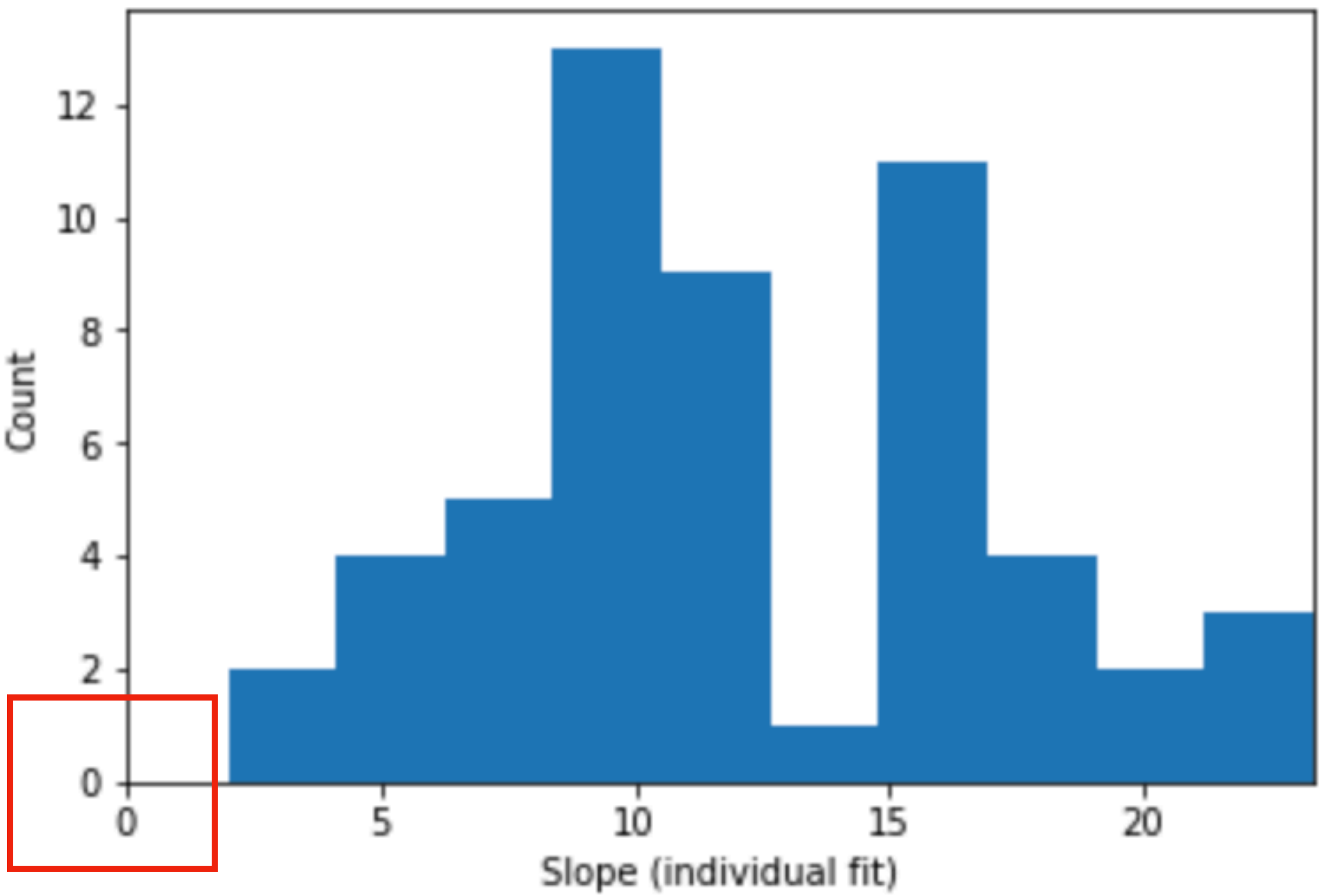
Lab 3: Mental rotation

Individual-level statistics also support the hypothesis



Lab 3: Mental rotation

Individual-level statistics also support the hypothesis



Null hypothesis: Histogram symmetric around 0

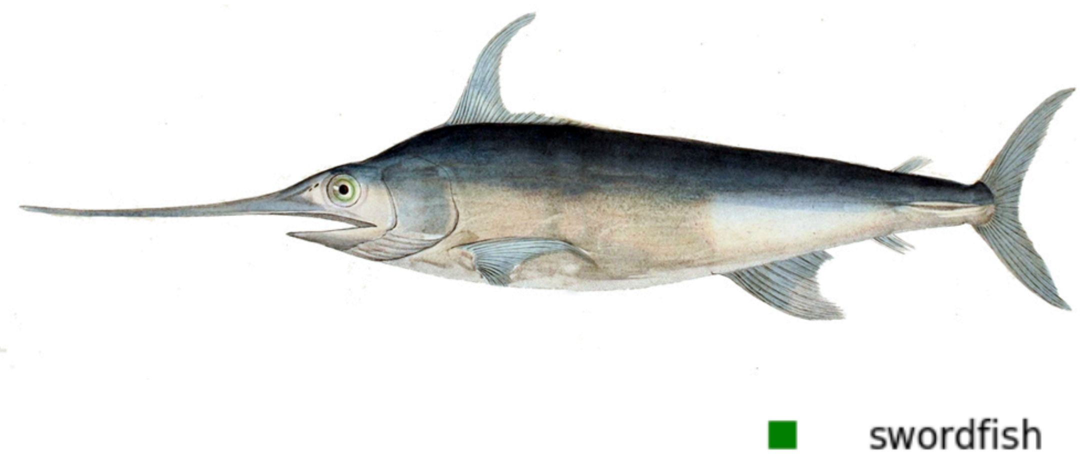
Outline

- How are categories formed?
  - Cognitive models of categorization
  - Machine classification
  - Model evaluation

Problem of categorization

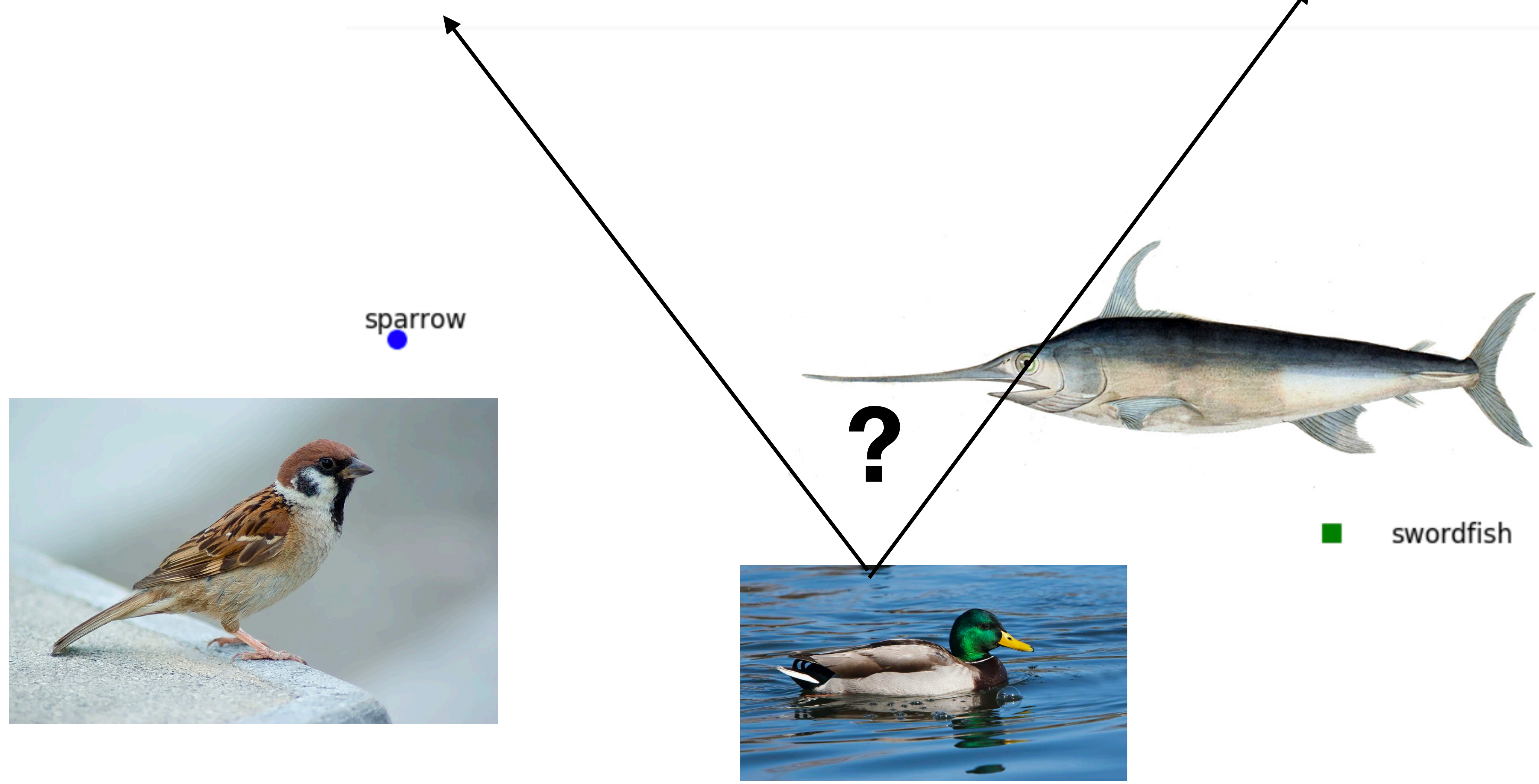
Category 1: Bird

Category 2: Fish



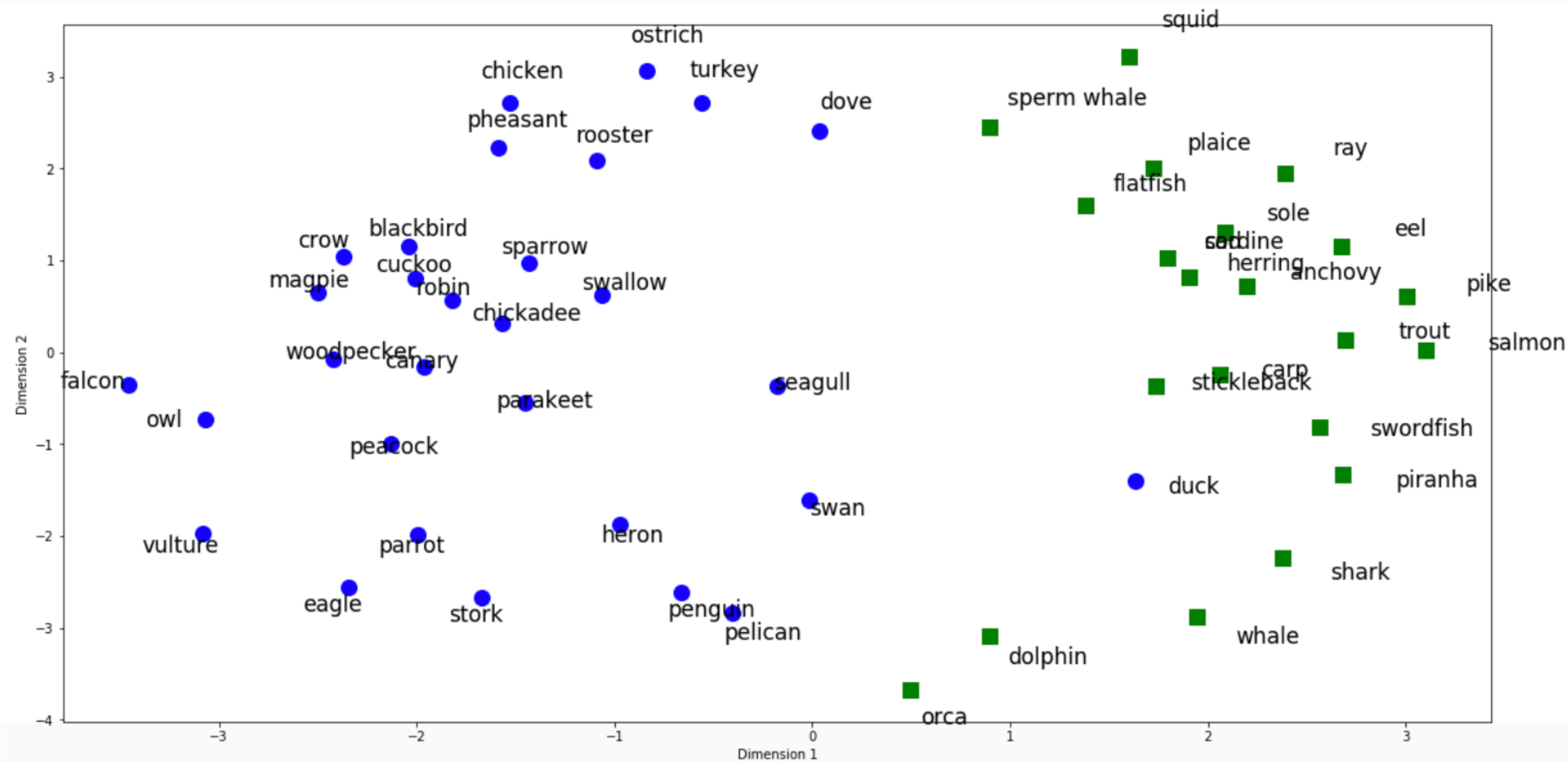
Problem of categorization

Category 1: Bird **Category 2: Fish**



Problem of categorization

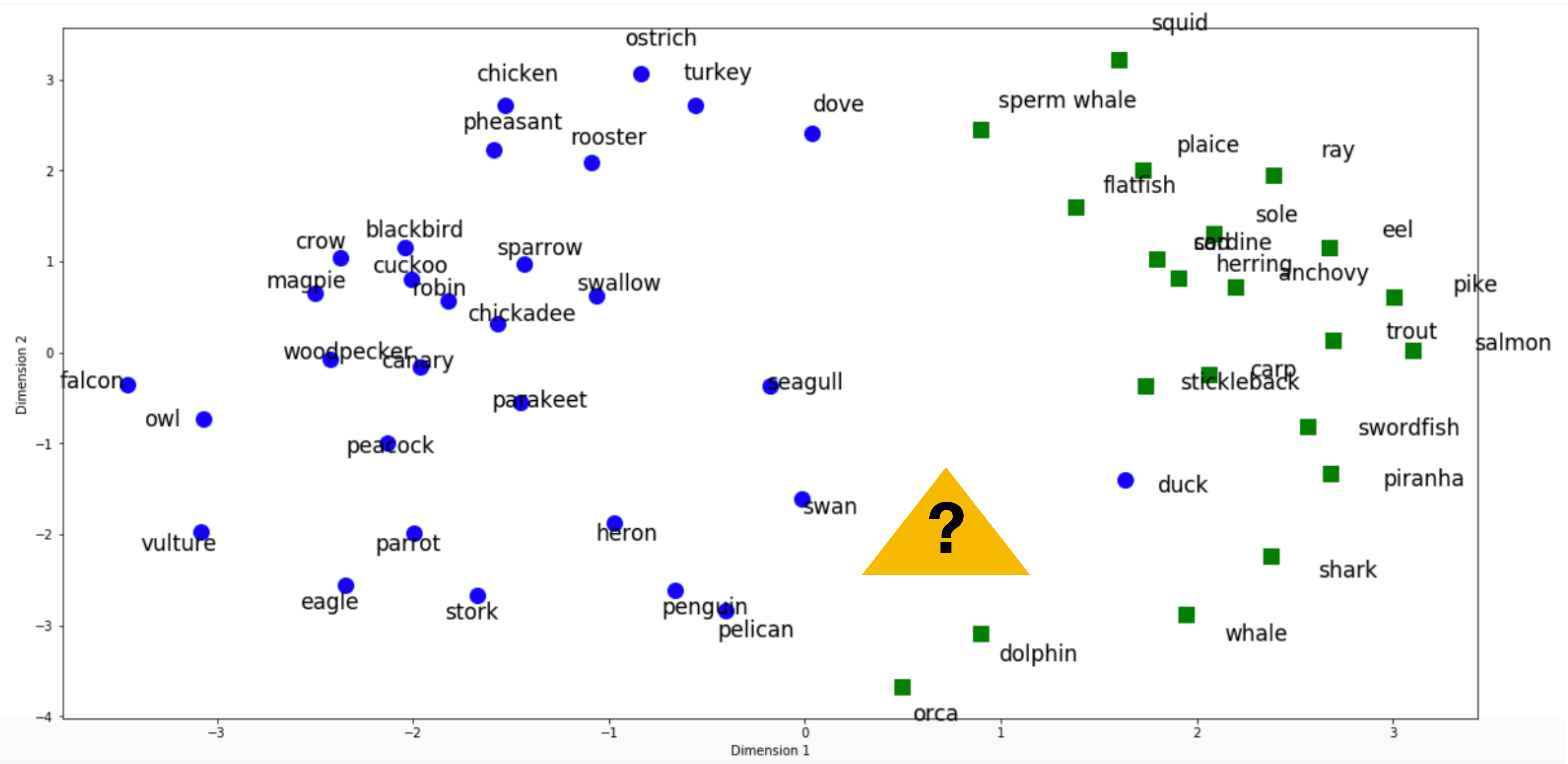
Category 1: Bird **Category 2: Fish**



Problem of categorization

Category 1: Bird

Category 2: Fish

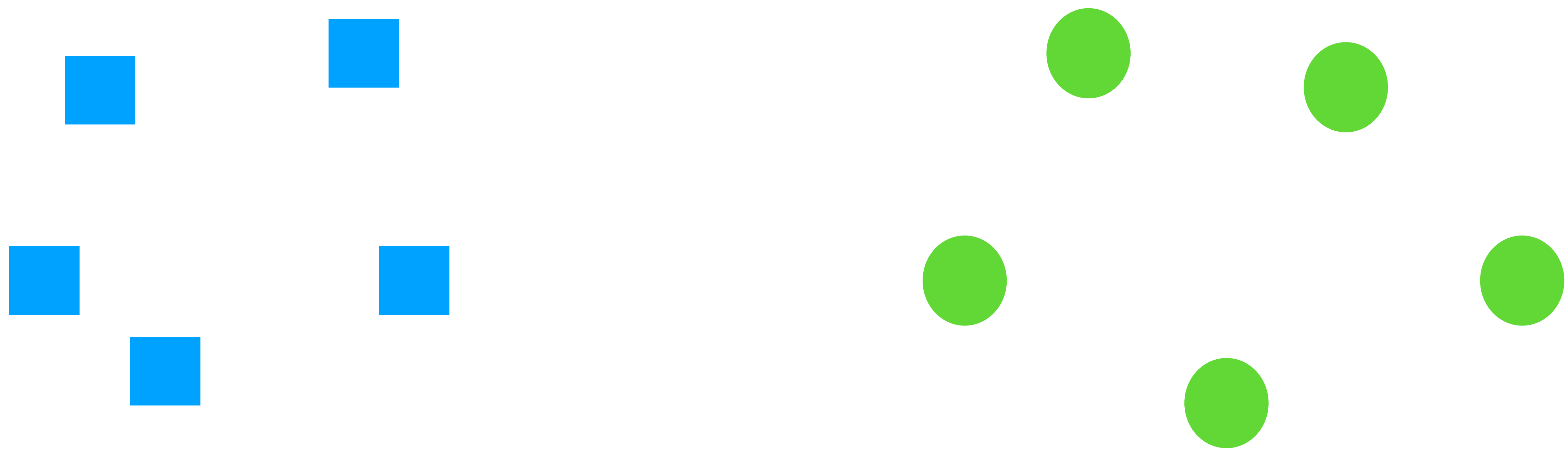


Cognitive models of categorization

- The prototype model (Reed, 1972; cf. Rosch, 1973)
- The exemplar model (Nosofsky, 1986)
- *Rational/Bayesian models (Anderson, 1991) [not discussed]

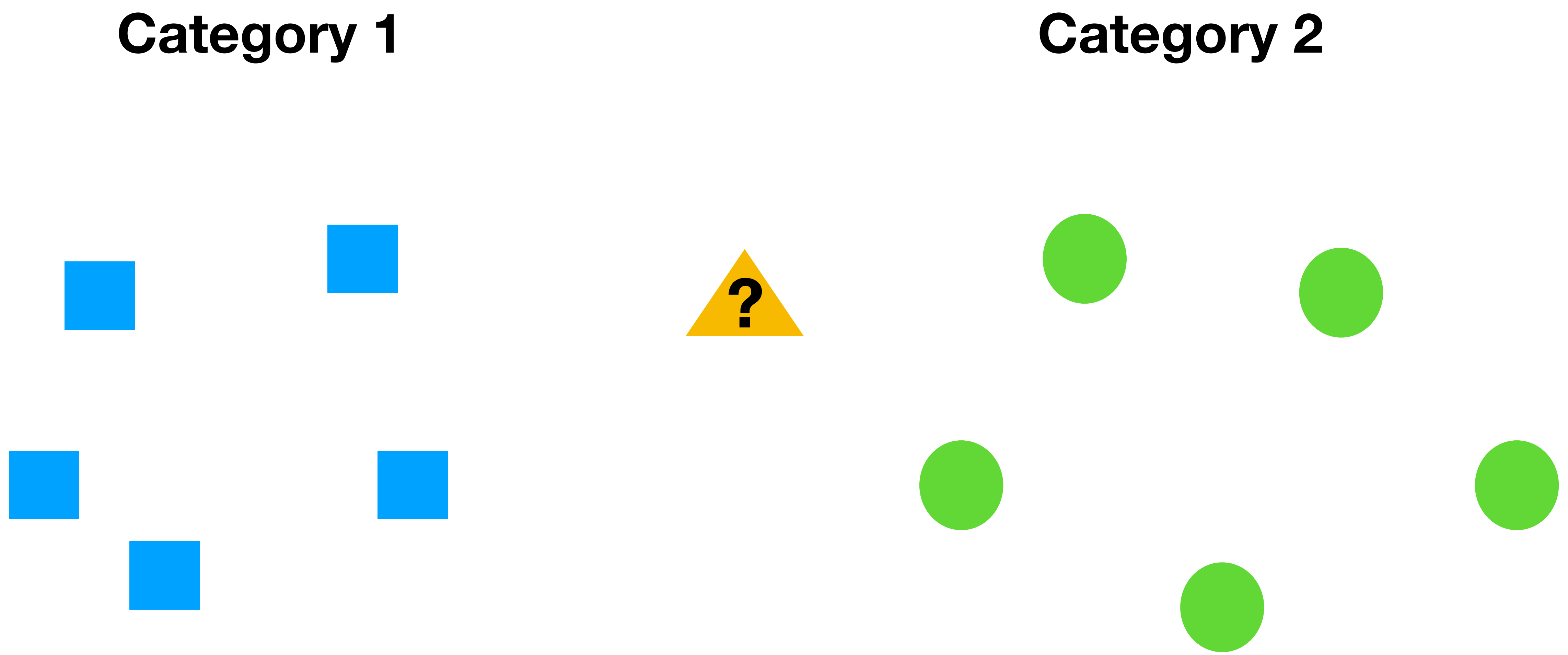
The prototype model

Category 1 **Category 2**



Reed (1972)

The prototype model

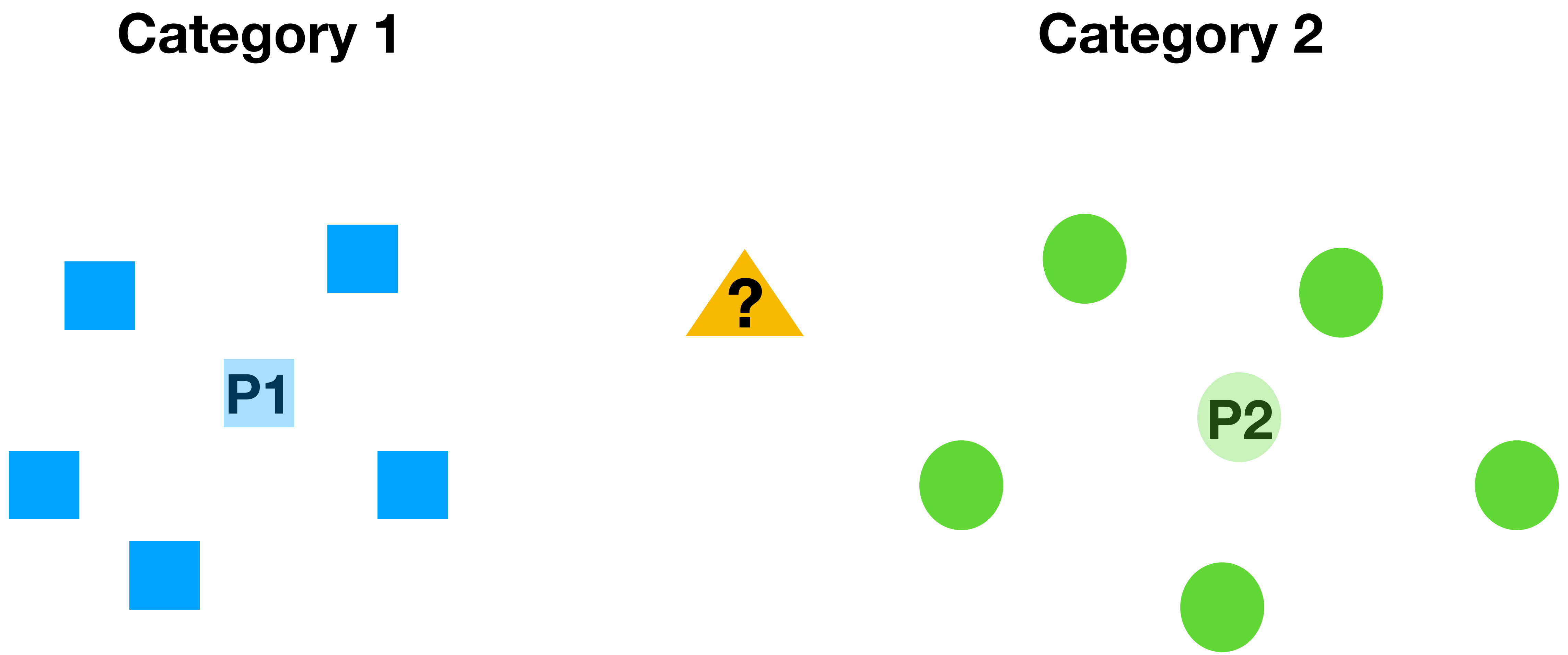


Reed (1972)

Discuss

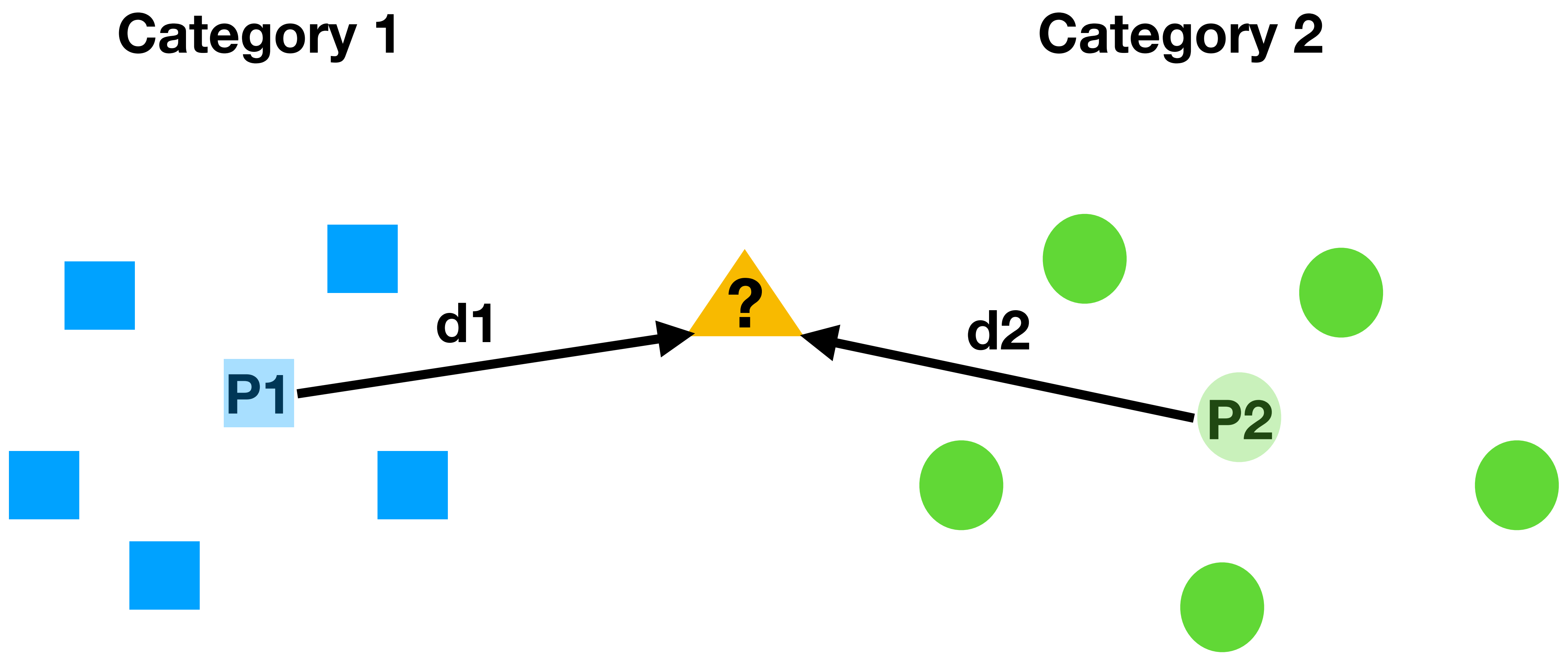
- Based on what we’ve learned from the prototype theory, how would you assign the category membership of a new item?

The prototype model



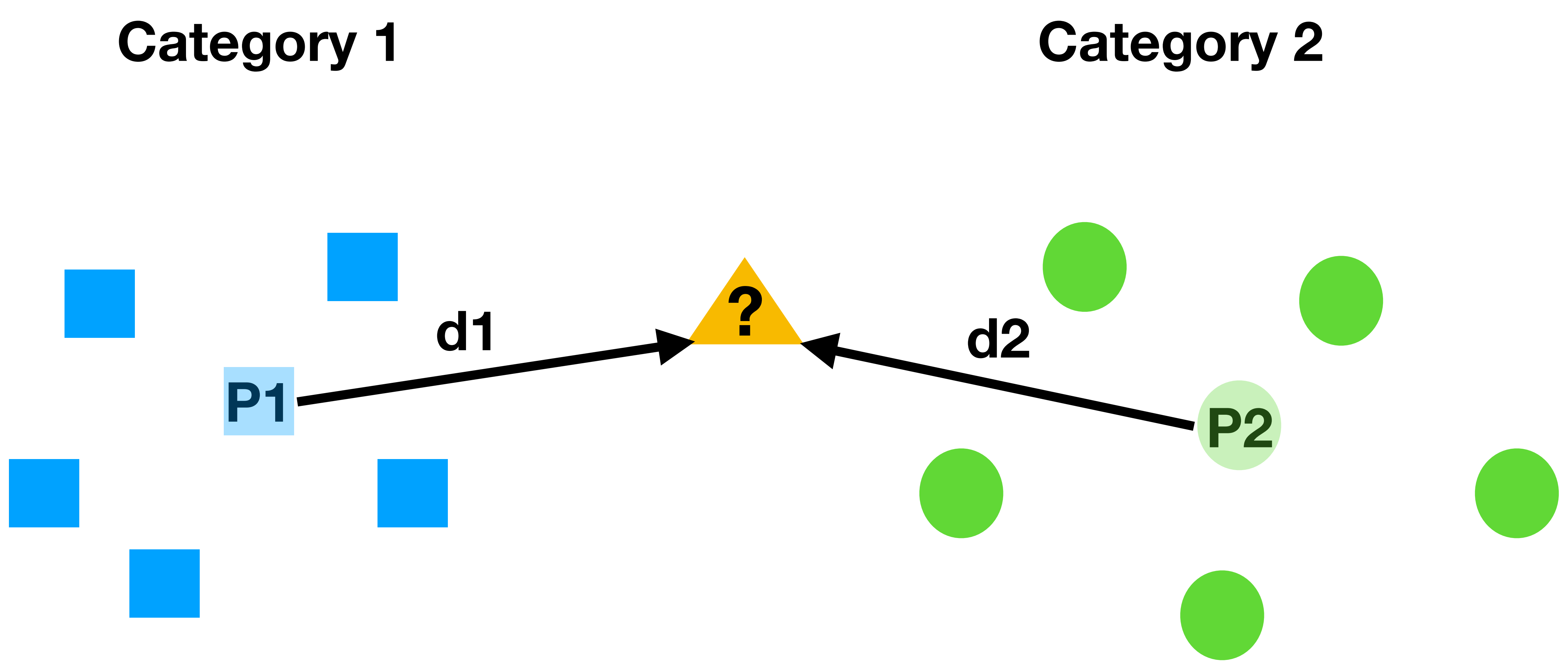
Reed (1972)

The prototype model



Reed (1972)

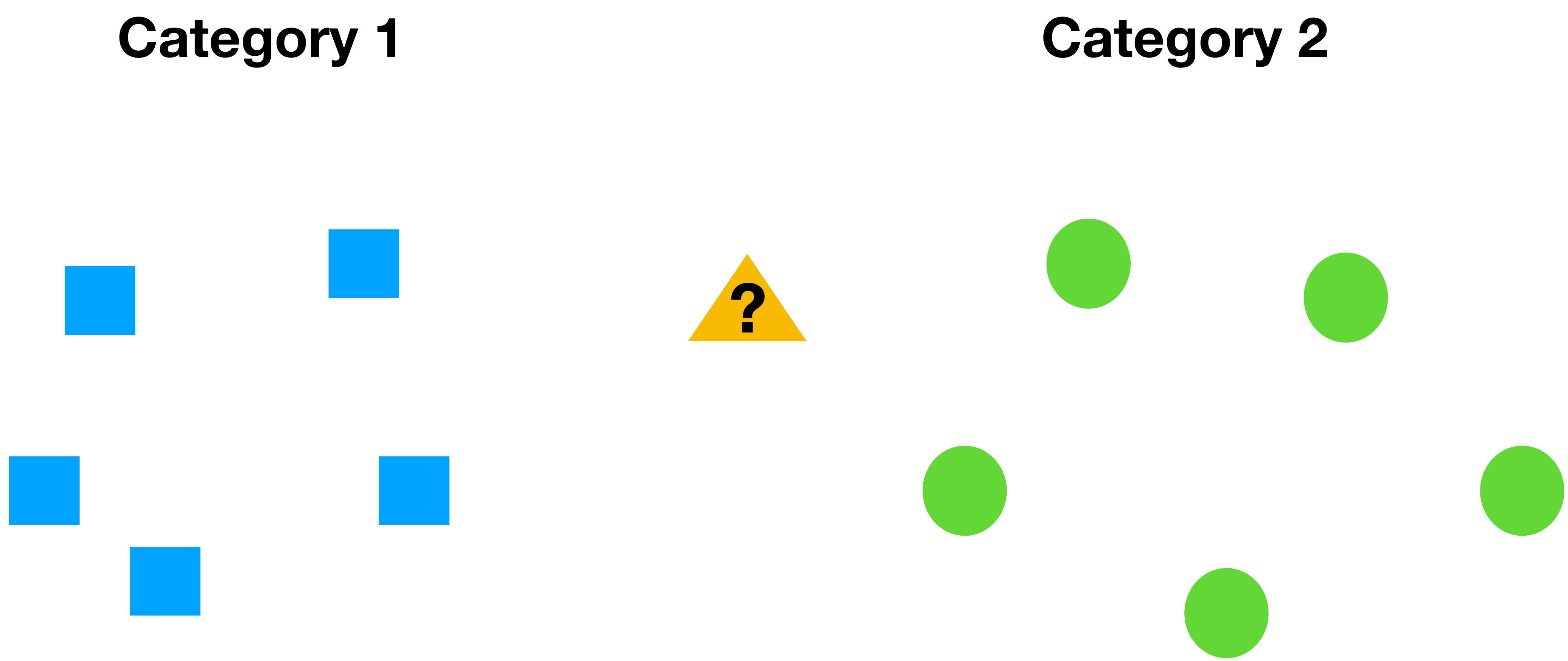
The prototype model



If d1 > d2, choose Category 2
Otherwise choose Category 1

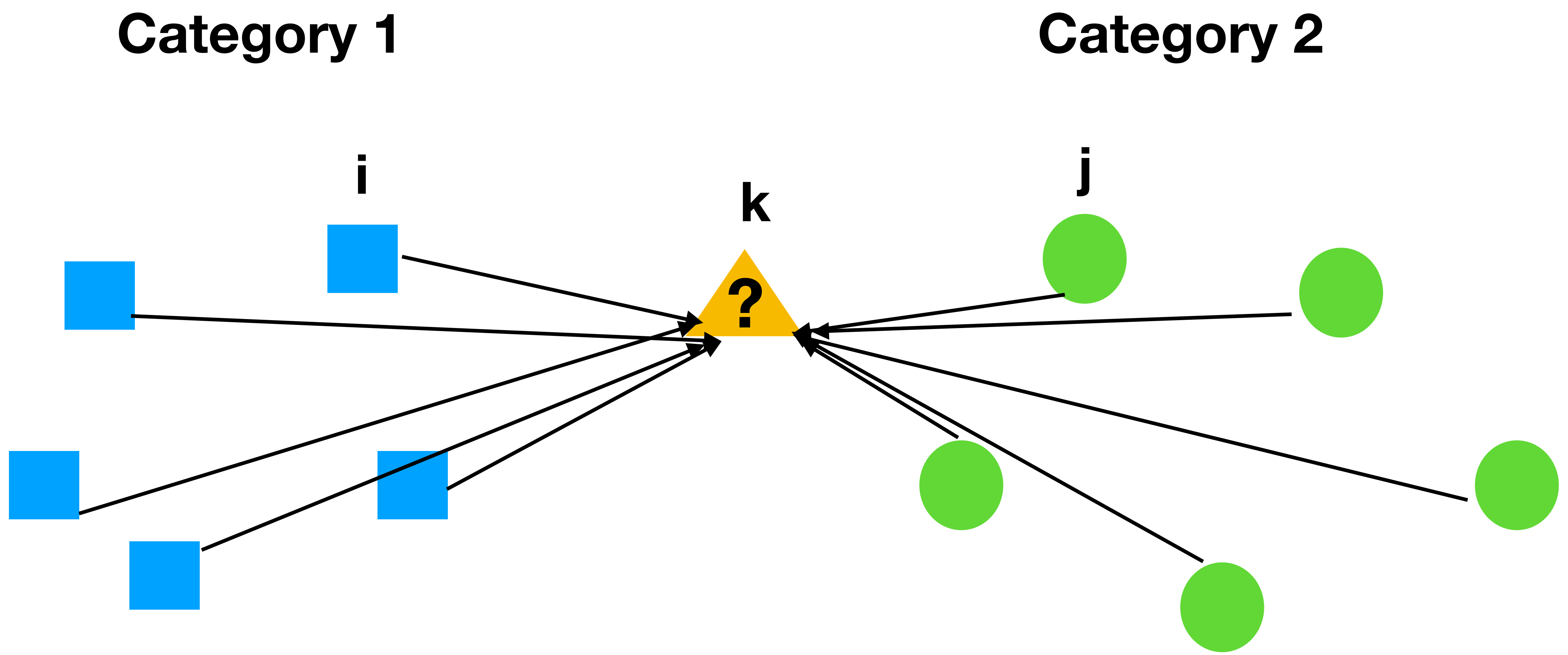
Reed (1972)

The exemplar model



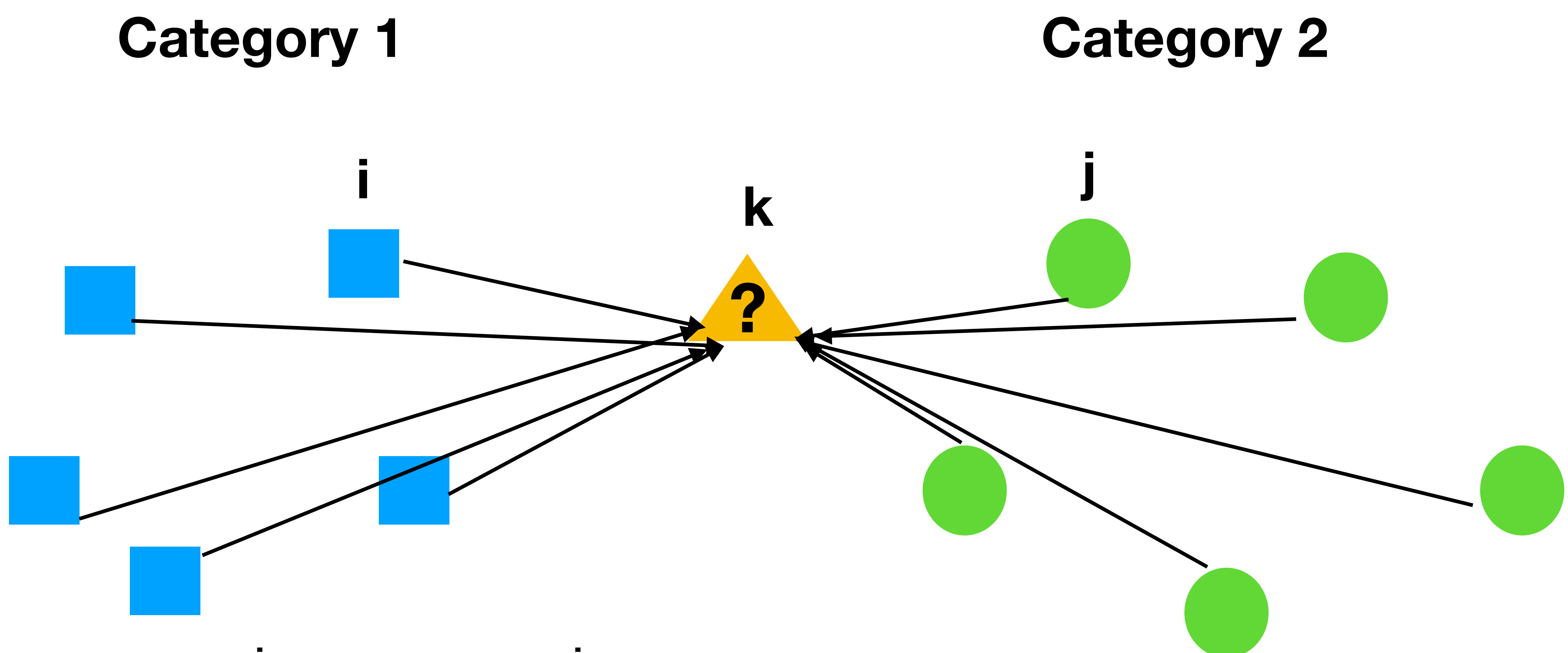
Nosofsky (1986)

The exemplar model



Nosofsky (1986)

The exemplar model



If $\frac{\sum^{i} \mathrm{sim}(i,k)}{\mathrm{Size(Cat1)}} > \frac{\sum^{j} \mathrm{sim}(j,k)}{\mathrm{Size(Cat2)}}$**, choose Category 1**

Otherwise choose Category 2

Nosofsky (1986)

Similarity function

$$sim(x,y) = \exp(-d(x,y)^2)$$

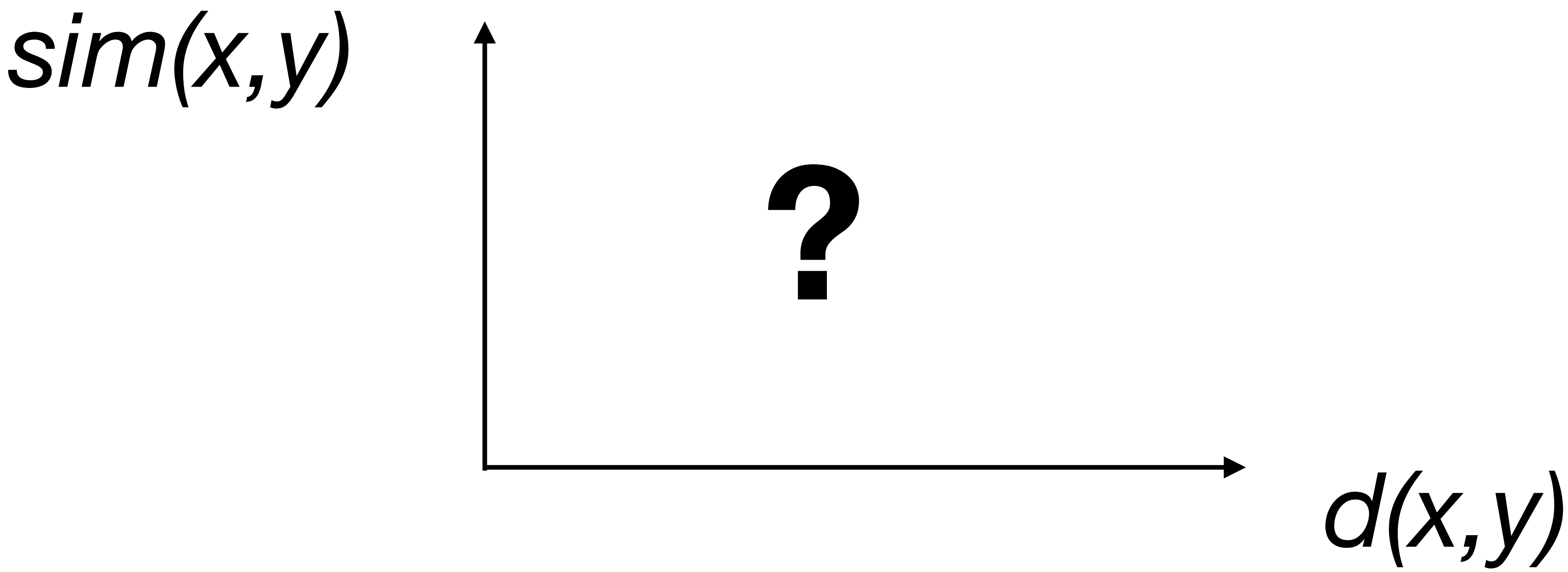
where $d(x,y)$ is Euclidean distance between x and y

Nosofsky (1986)

Similarity function

$$sim(x,y) = exp(-d(x,y)^2)$$

where $d(x,y)$ is Euclidean distance between x and y



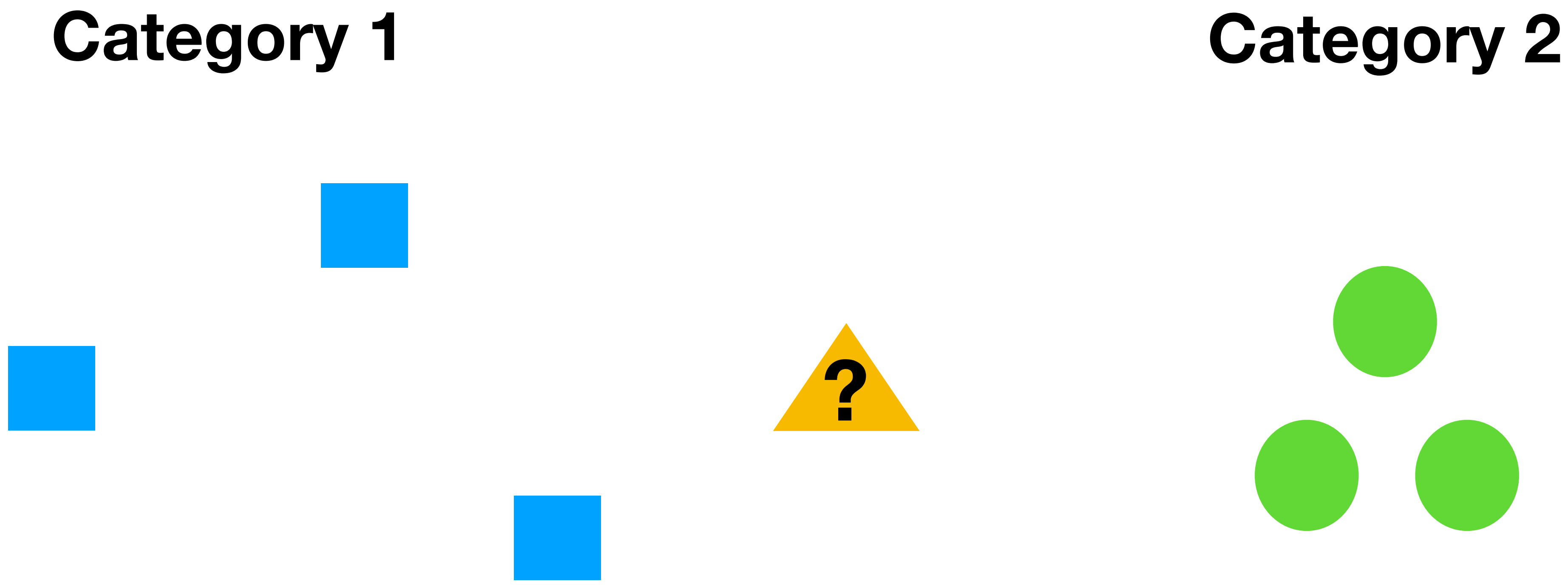
Nosofsky (1986)

Discuss

- When do prototype and exemplar models make *different* predictions?

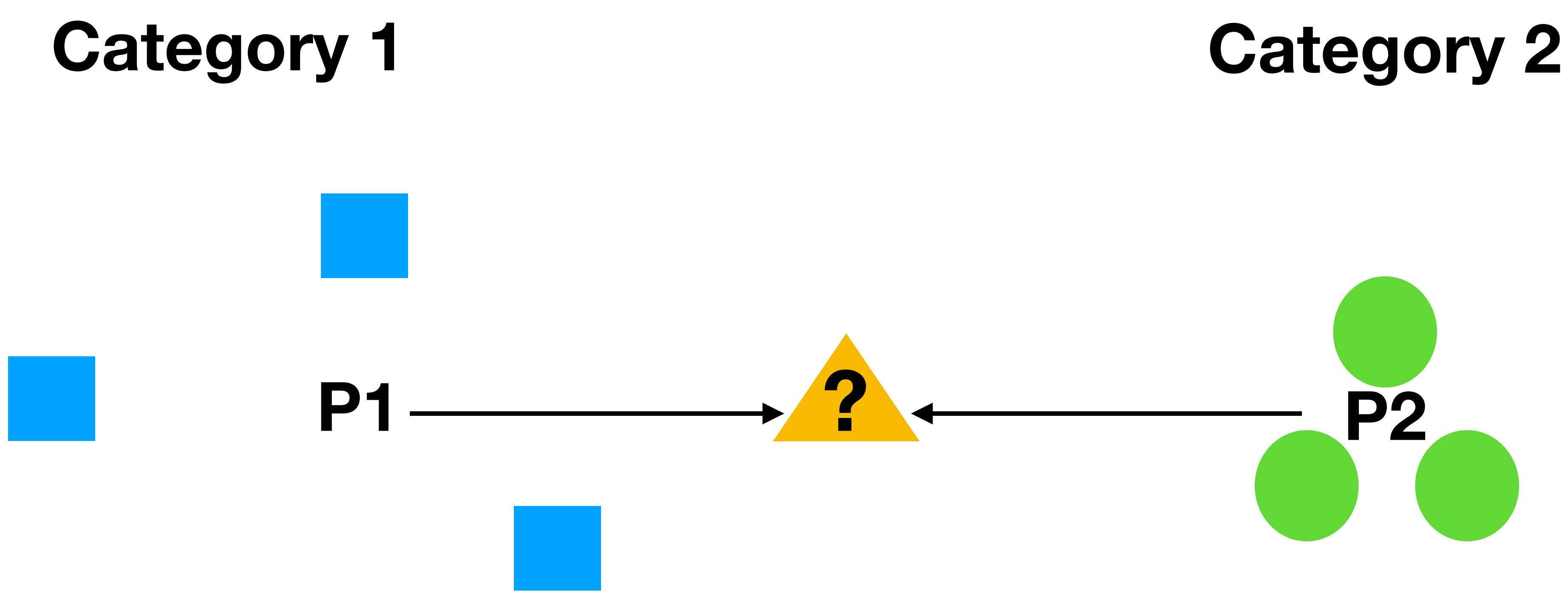
Discuss

- Consider the following scenario:



Prototype prediction

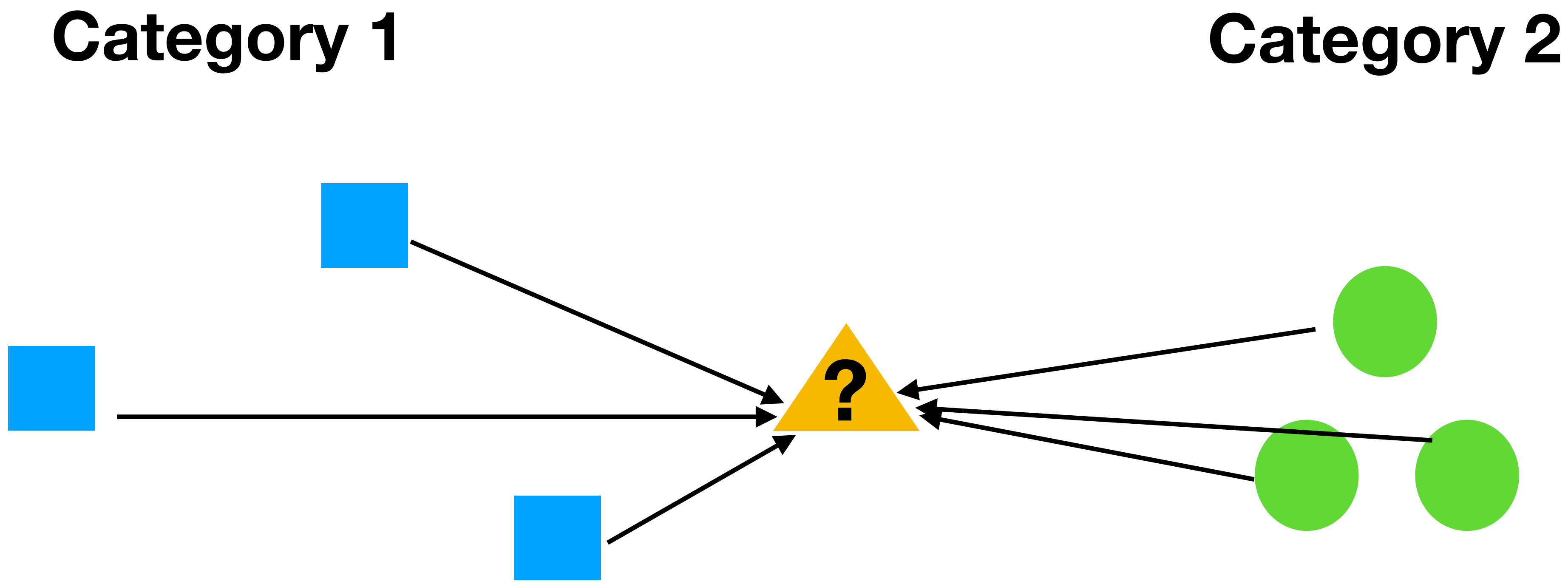
- Category prototypes are equidistant to a query.



Prototype model would predict either category

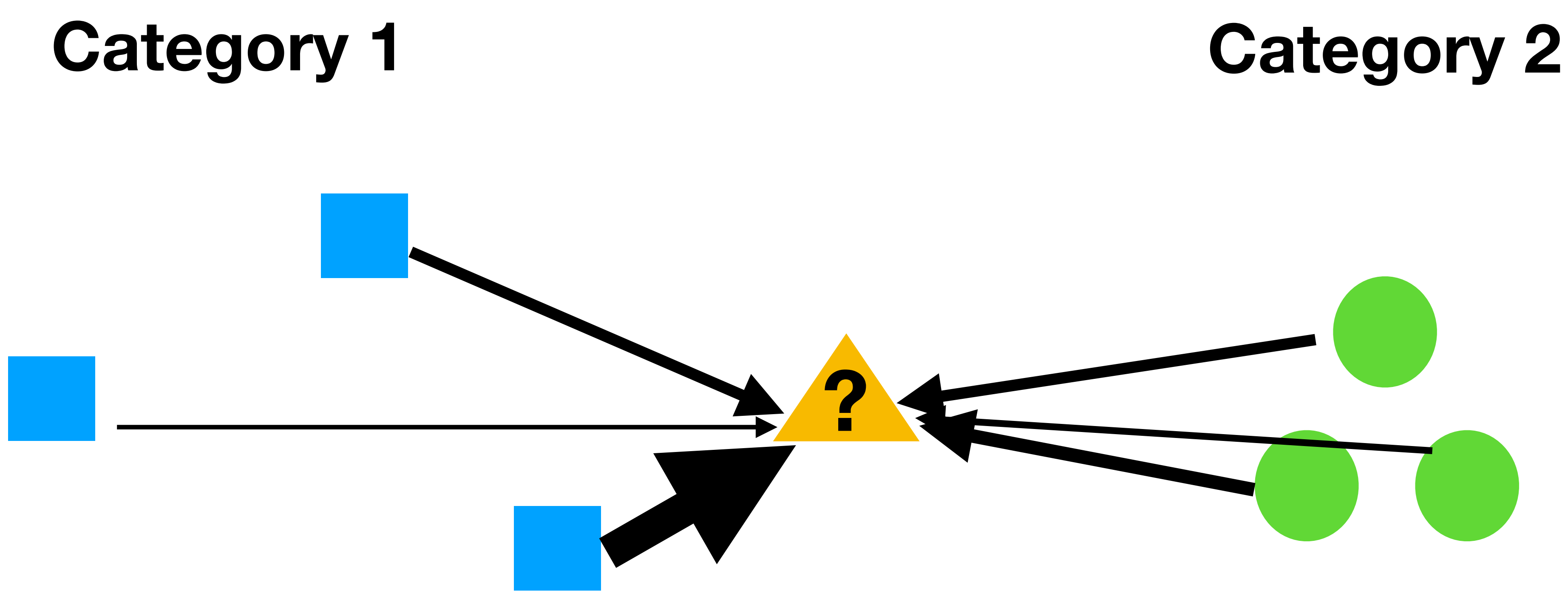
Exemplar prediction

- Exemplar model considers influence from each instance or exemplar to a query.



Exemplar prediction

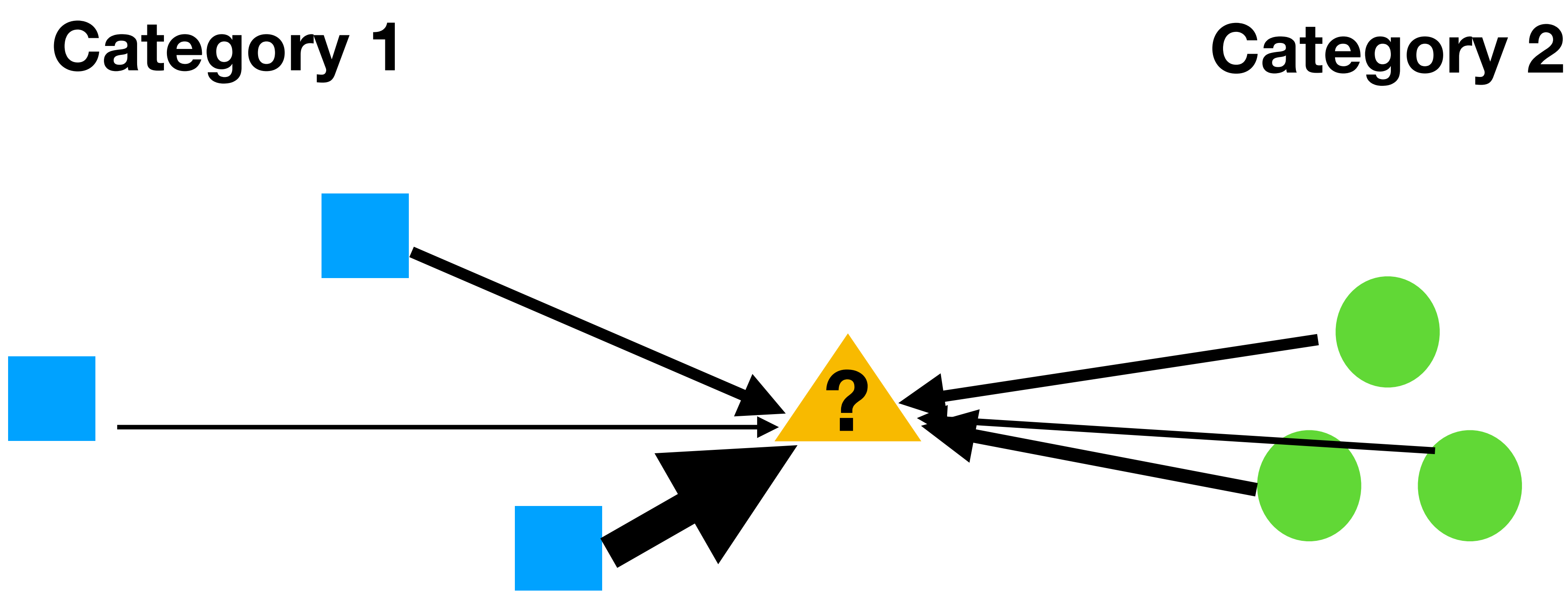
- Except that the distance measure is transformed non-linearly, effectively weighting the exemplars.



Exponential transform captures neighbourhood density

Exemplar prediction

- While the prototype model predicts based on averages, the exemplar model captures density.



Exponential transform captures neighbourhood density

Exemplar model would more likely predict Category 1

5-minute break

Machine classification

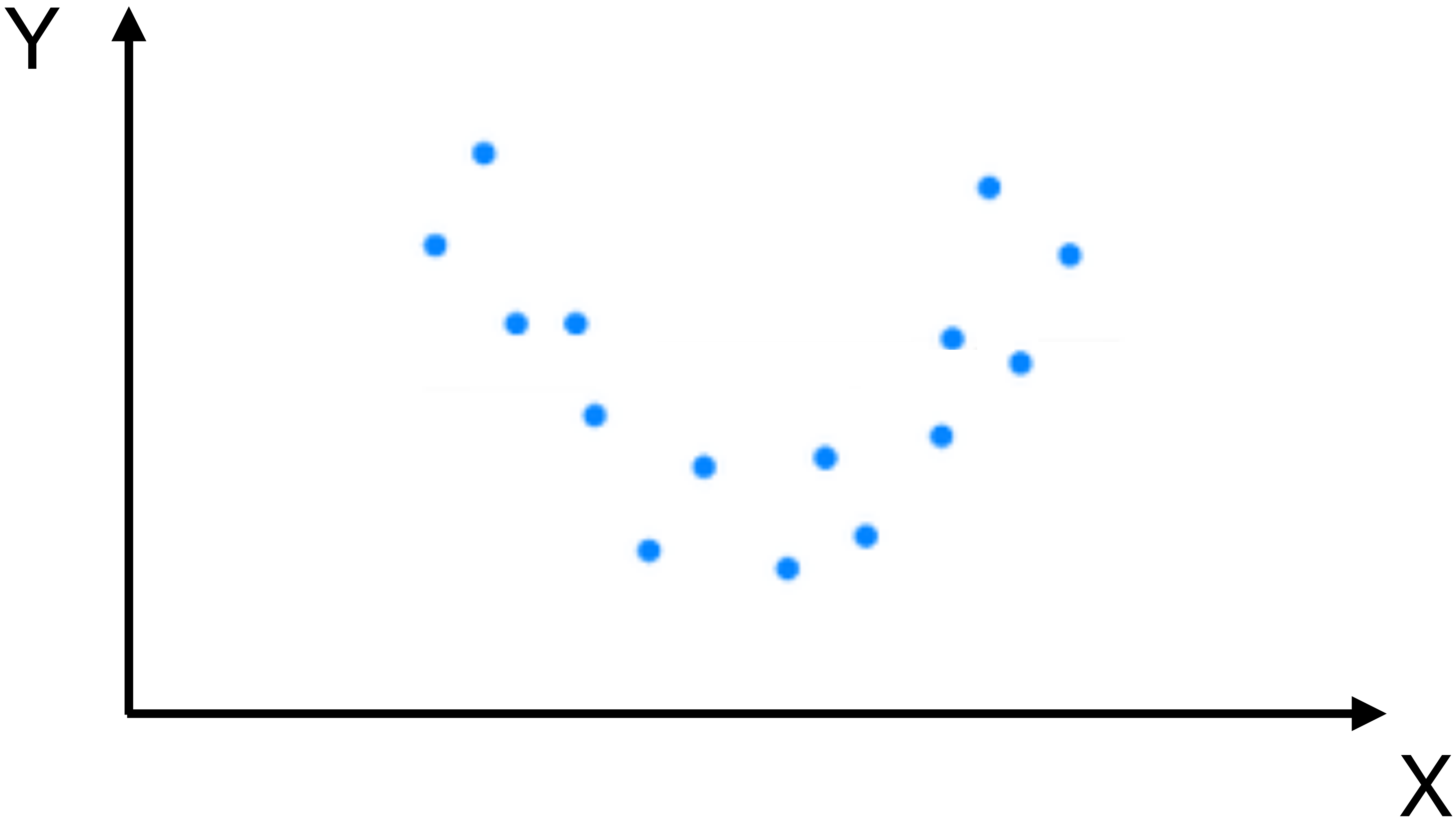
- Supervised machine learning (regression, classification)
- Gaussian and *k*NN classifiers
- Overfitting and cross validation

Supervised machine learning

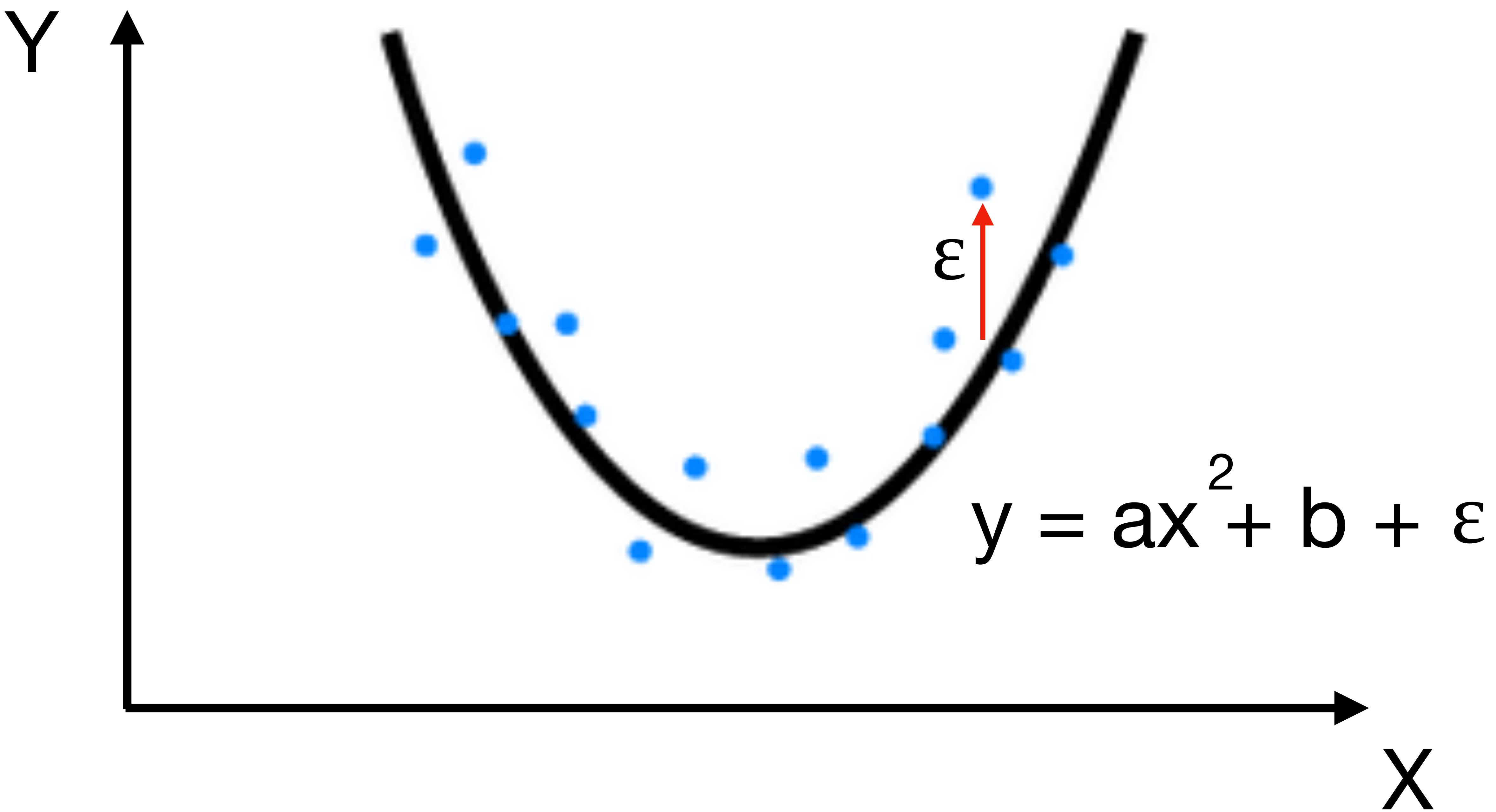
$$Y = f(X) + \varepsilon$$

- Y: Output variables to be predicted, or dependent variables
- X: Input variables, or independent variables
- $f(.)$: Function that maps input to output
- ε: Error, or uncertainty

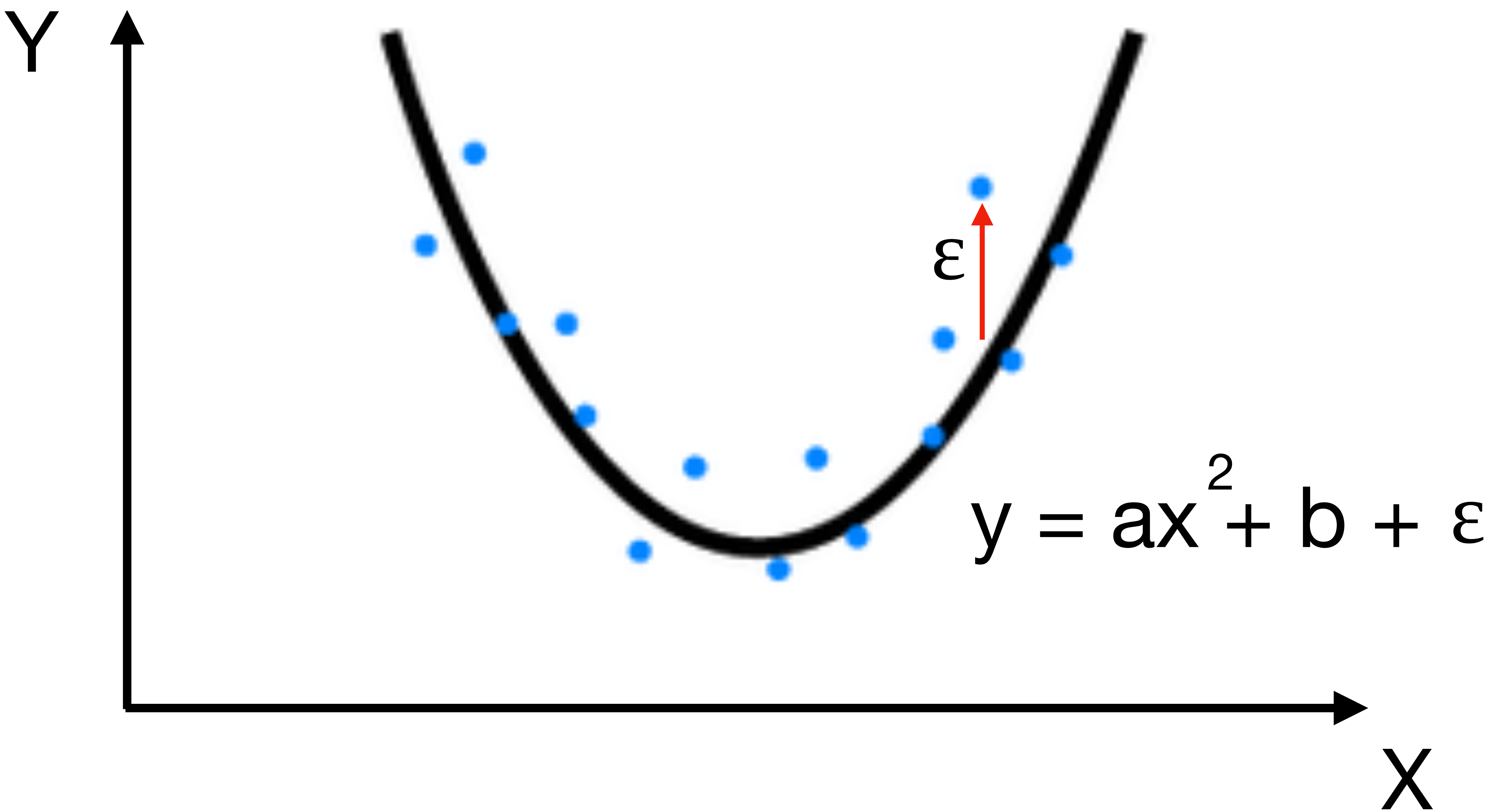
Regression



Regression



Regression



e.g. financial prediction (stock price)

Classification

X

Y

$f(.)$

—>

Flower

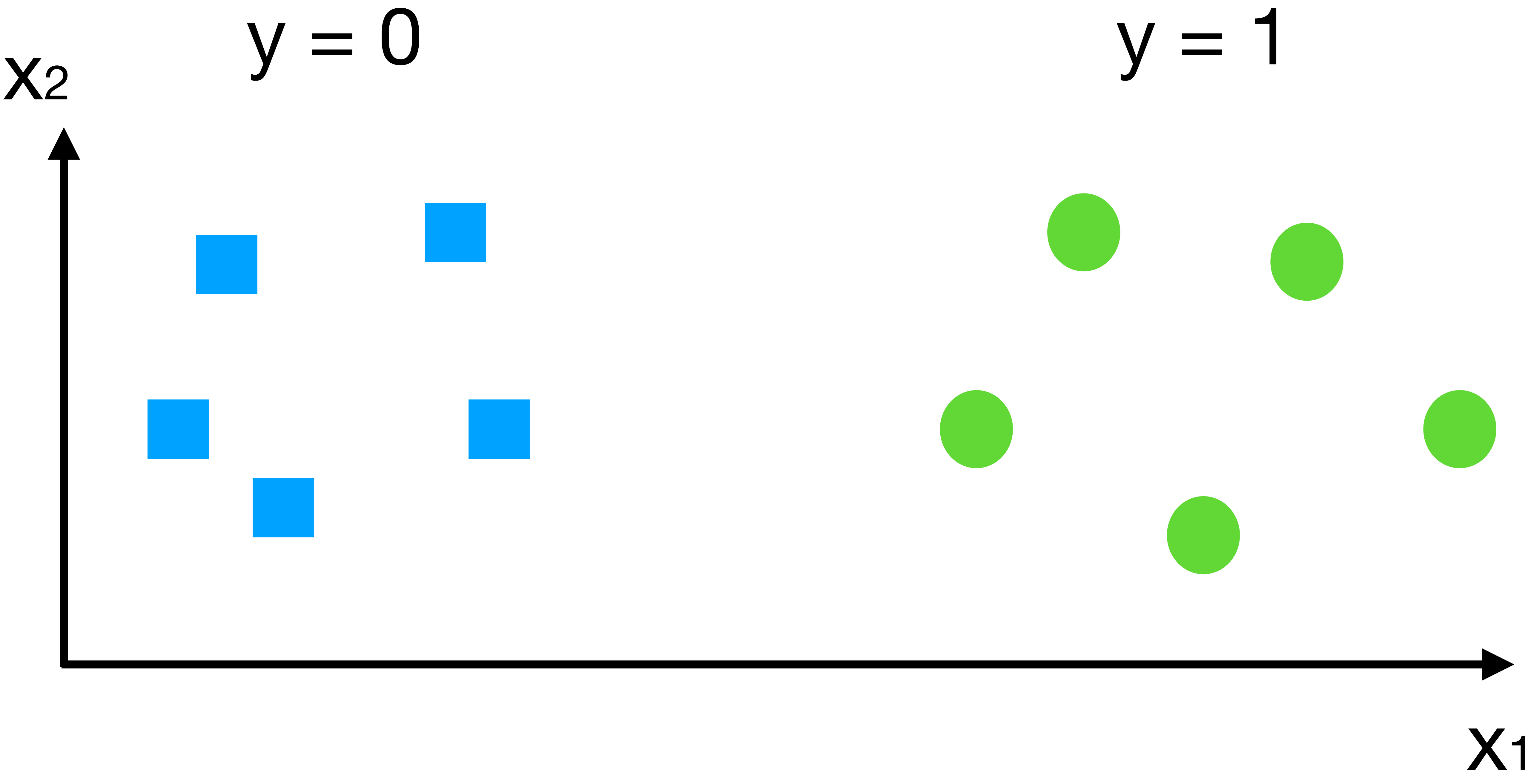
Elephant

Submarine

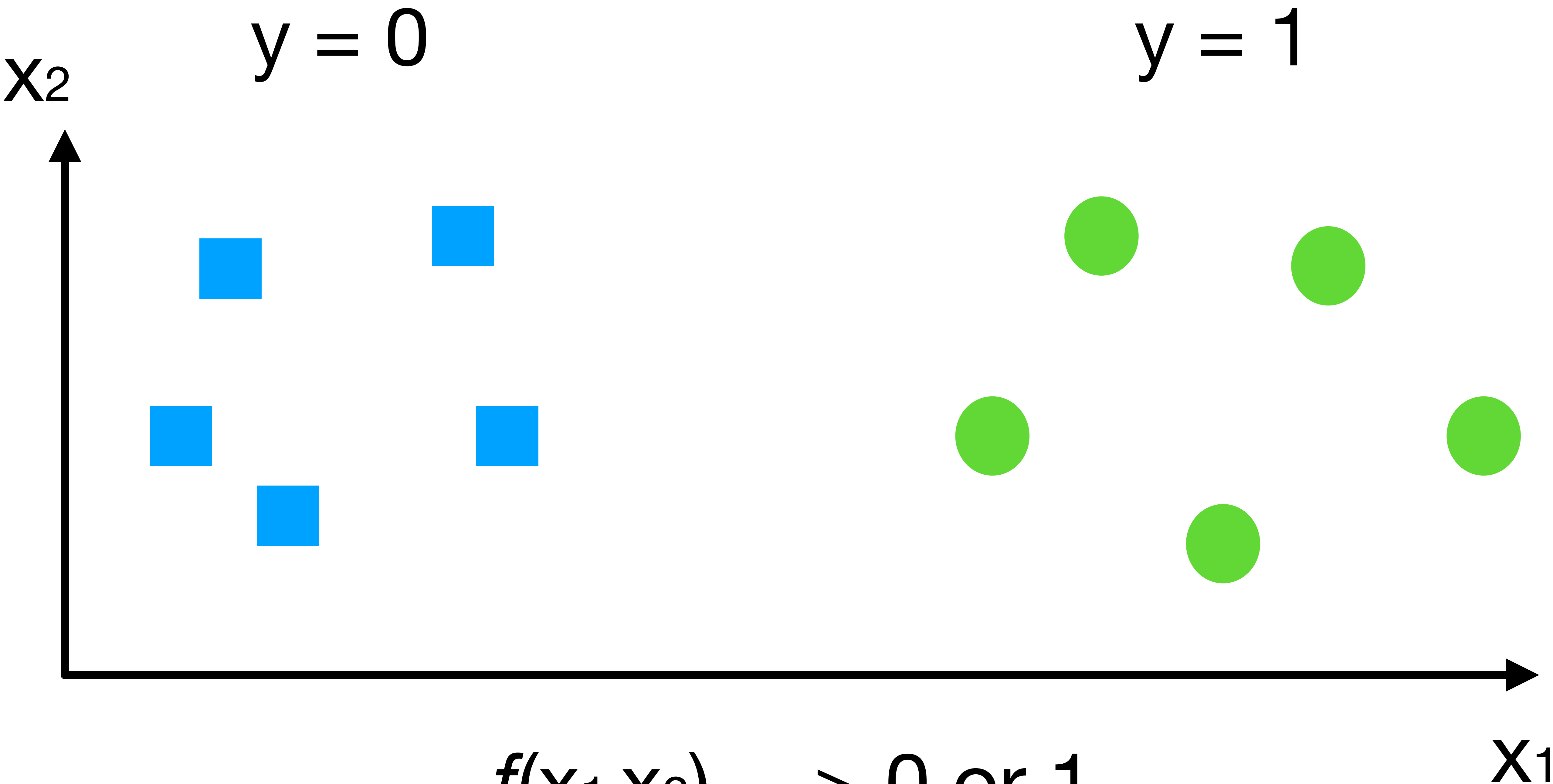
Pumpkin

Dog

Classification

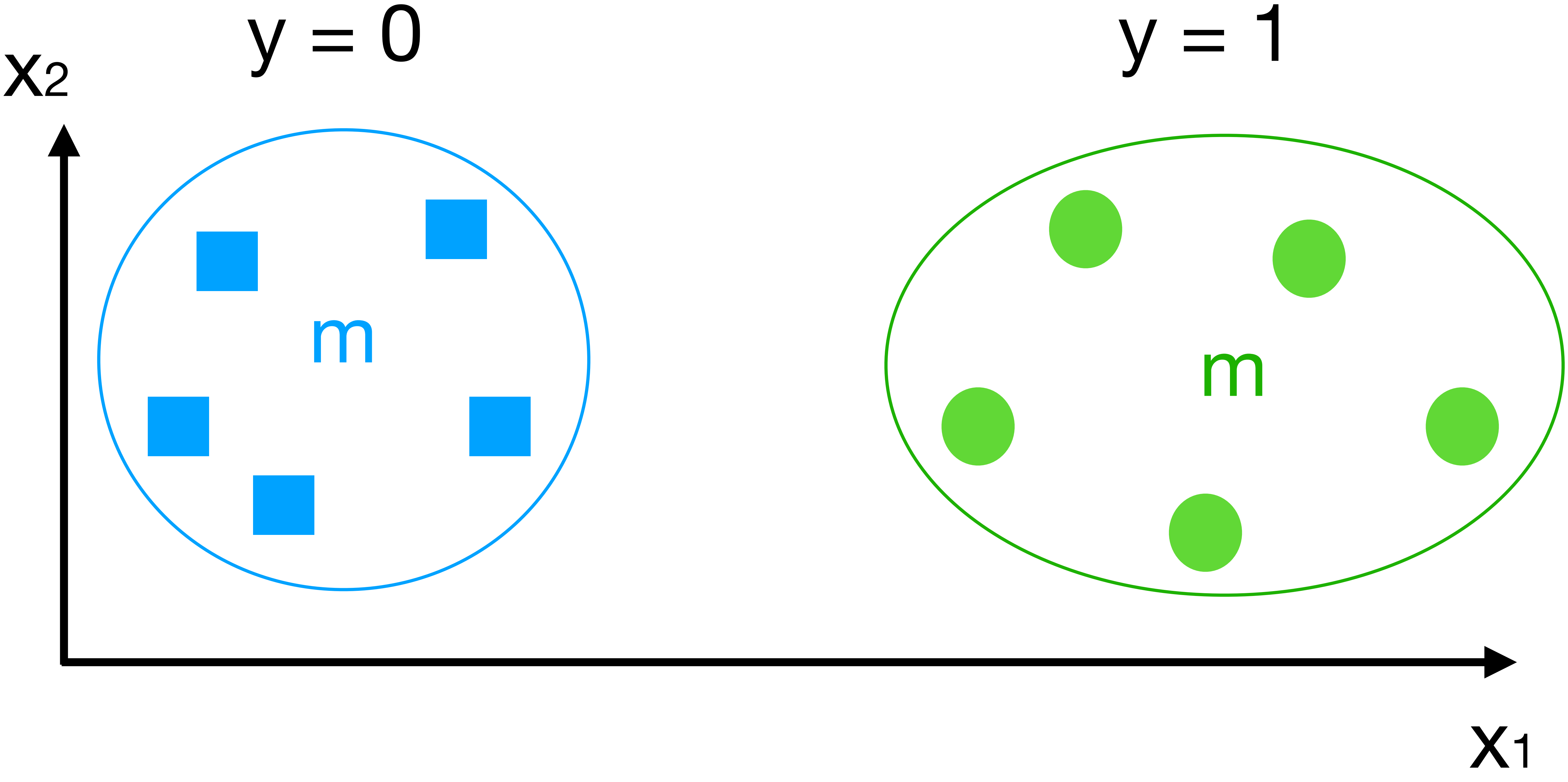


Classification

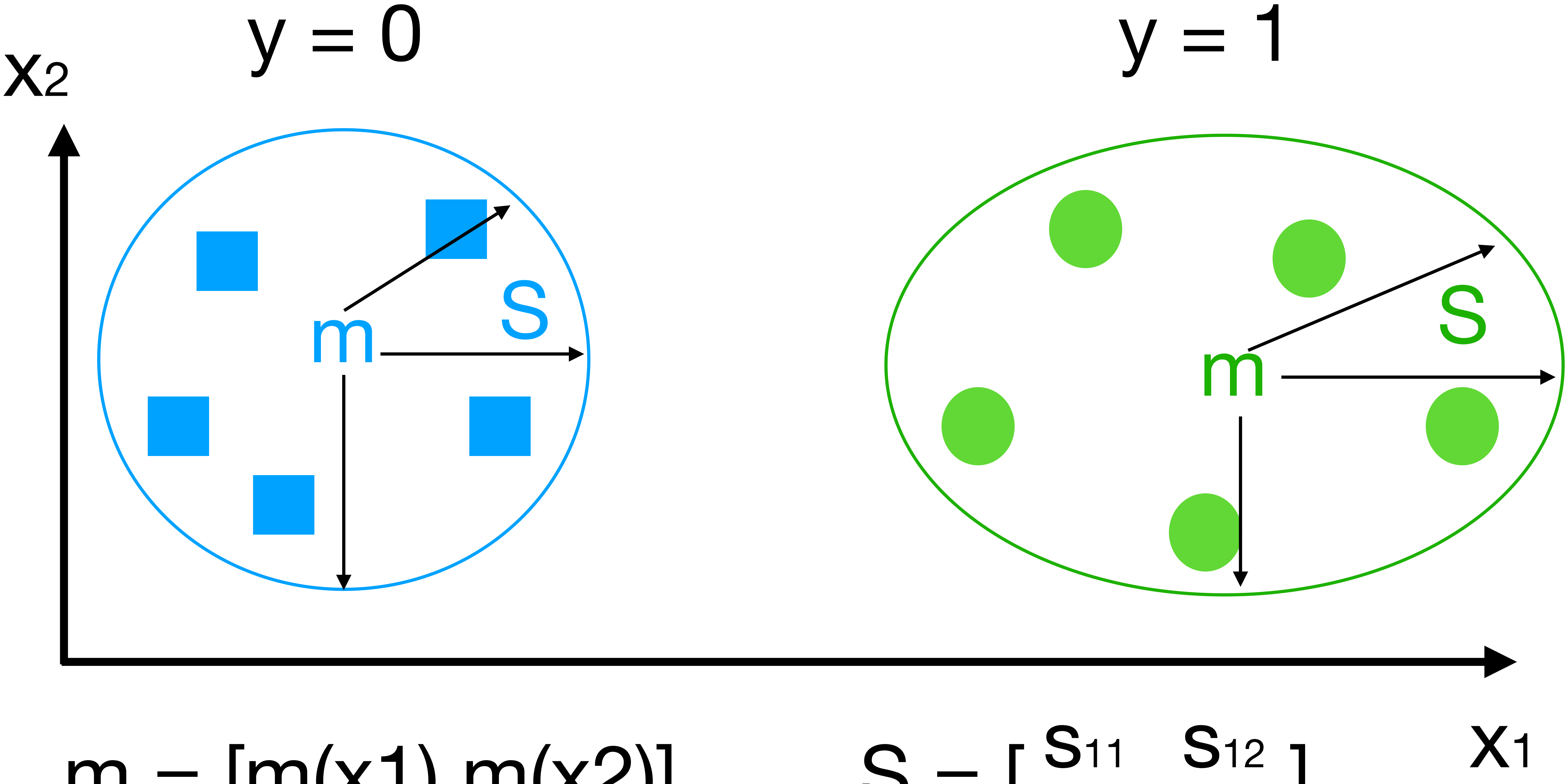


$f(x_1,x_2) \longrightarrow 0$ or 1

Gaussian classifier

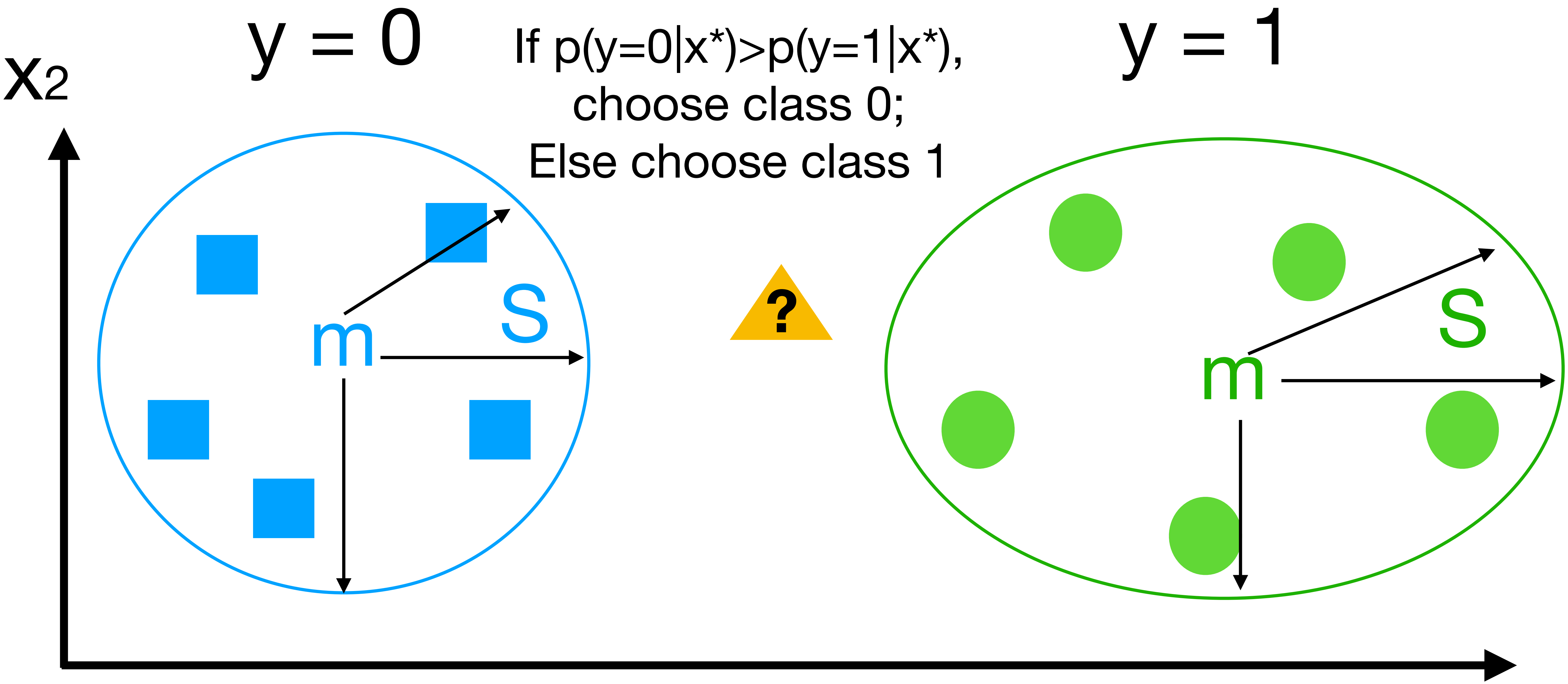


Gaussian classifier



$$m = [m(x1), m(x2)] \qquad S = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$$

Gaussian classifier



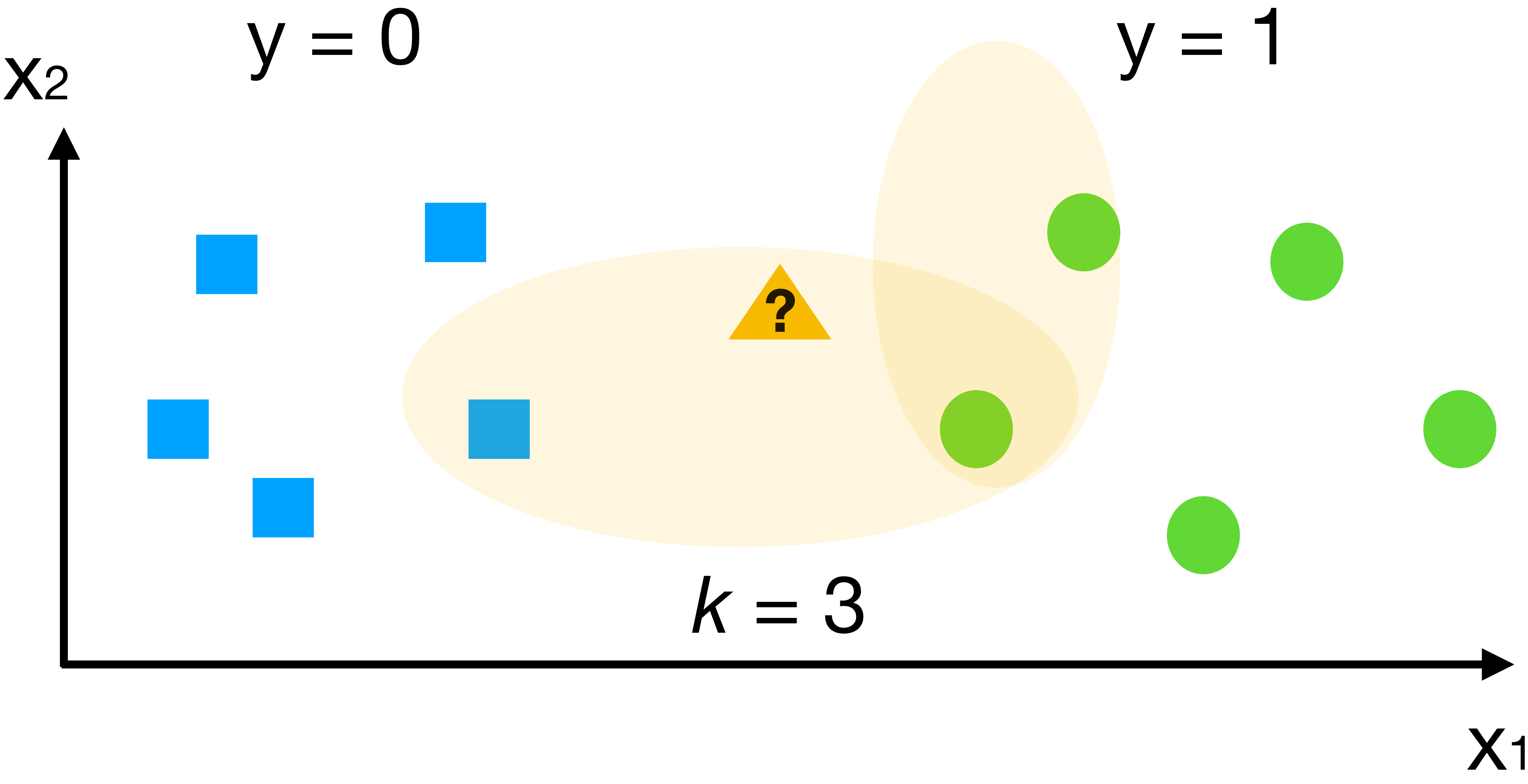
$m = [m(x1), m(x2)]$

$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$

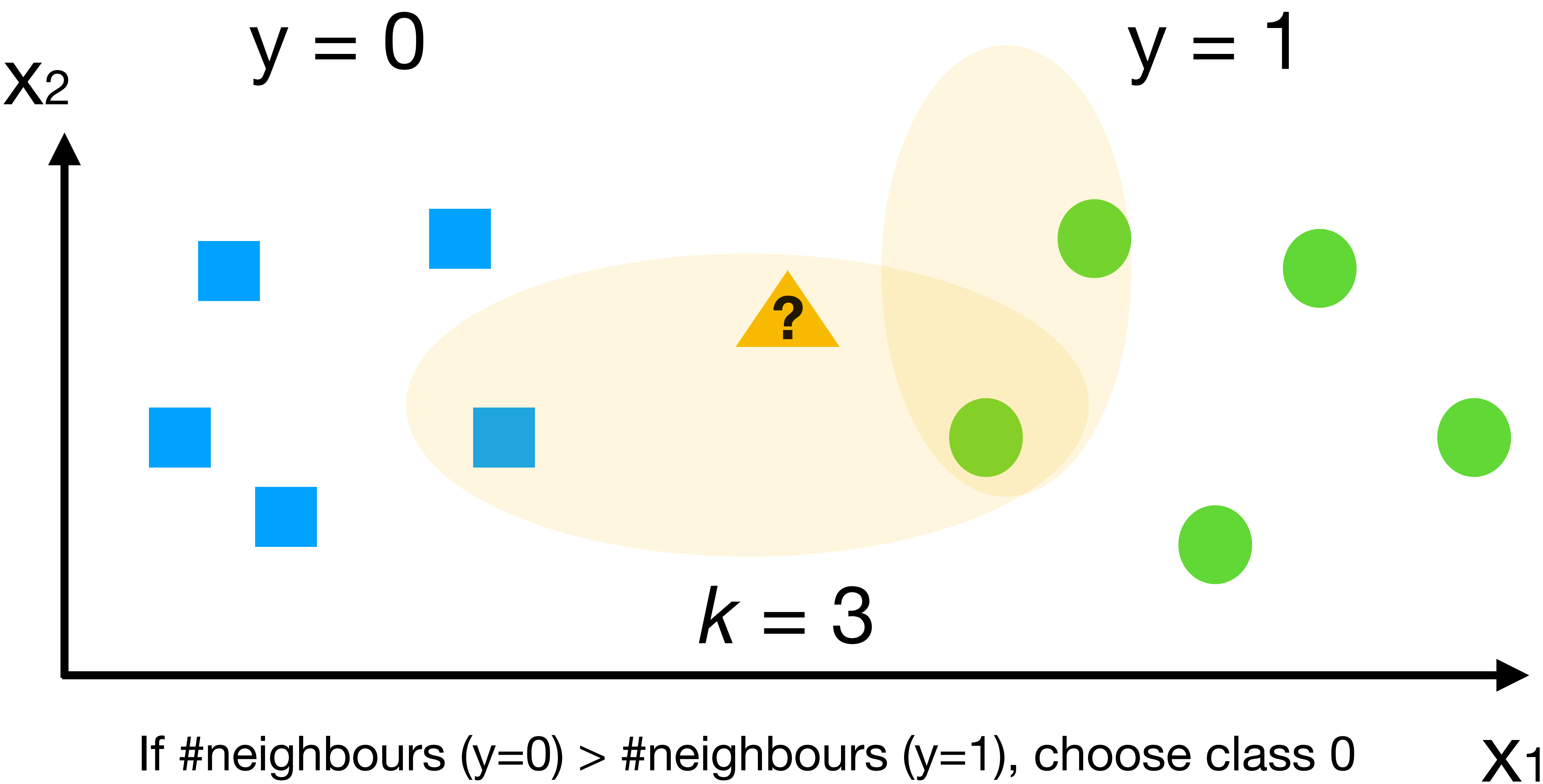
Connection with the prototype model

- The prototype model may be considered as a special case of the Gaussian classifier, namely:
  - The prototype model only takes into account means
  - No variances, (or that categories have same variances)
- As a result, the Gaussian classifier is more flexible than the prototype model (discuss why, e.g. in 1D)

k-nearest-neighbour classifier

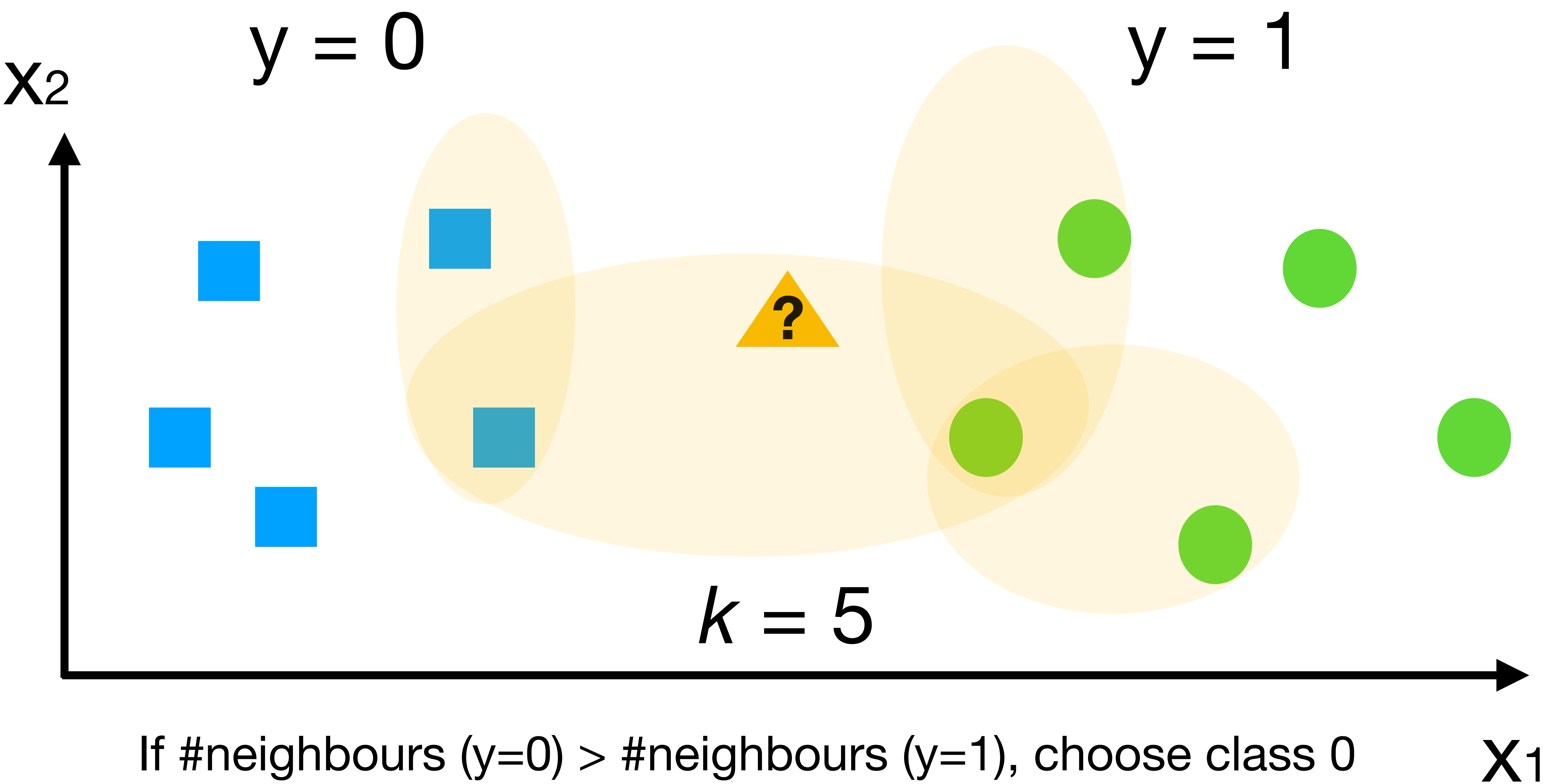


k-nearest-neighbour classifier



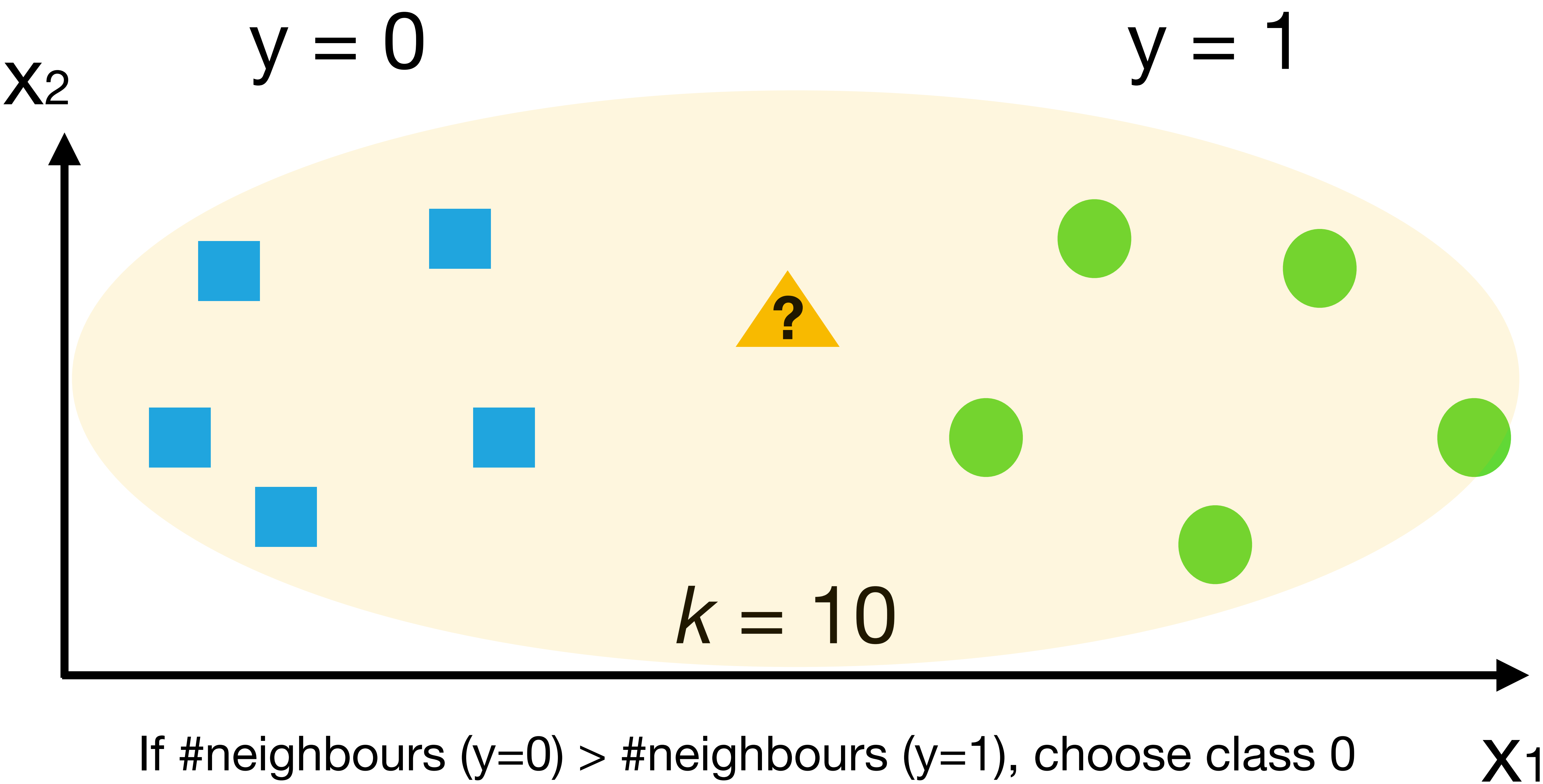
If #neighbours (y=0) > #neighbours (y=1), choose class 0
Else choose class 1

k-nearest-neighbour classifier



If #neighbours (y=0) > #neighbours (y=1), choose class 0
Else choose class 1

k-nearest-neighbour classifier



If #neighbours (y=0) > #neighbours (y=1), choose class 0
Else choose class 1

Connection with the exemplar model

- The exemplar model is similar to the *k*NN classifier except:
  - The actual distances matter in exemplar model
  - Only distance ranks matter in *k*NN
  - *k*NN explicitly selects the pool of exemplars for classification (by choice of *k*), and exemplar model takes into account all exemplars (i.e., soft/weighted choice)

Summary

Prototype model	Classify by measuring distances to prototypes	Similar to a Gaussian classifier but with no variances
Exemplar model	Classify by measuring average similarities of all exemplars	Similar to a soft and extreme version of k-nearest-neighbour classifier

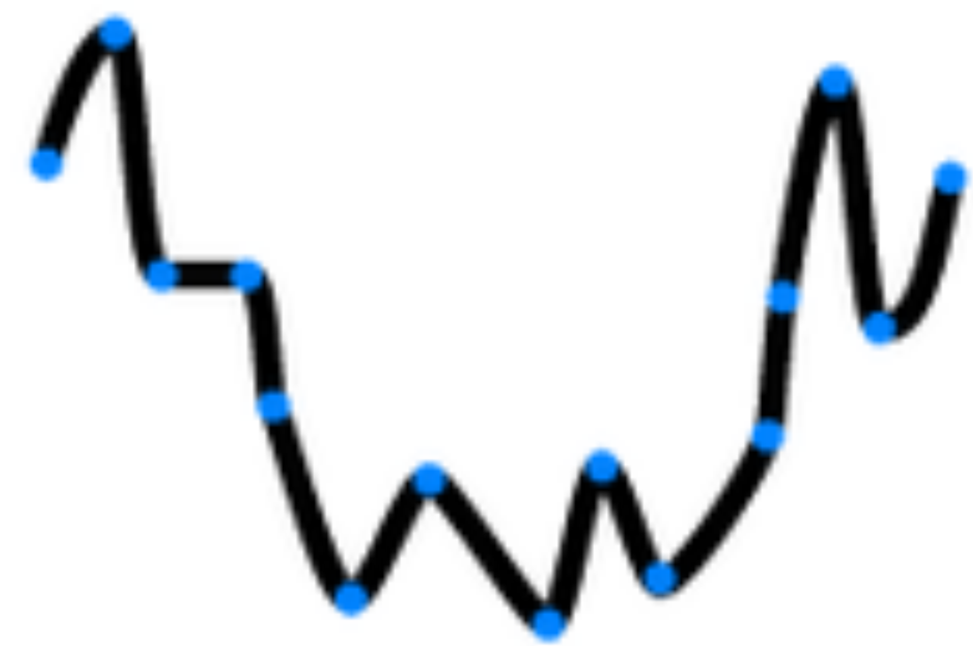
5-minute break

The risk of overfitting

$$y = ax^2 + bx + c + \varepsilon$$

The risk of overfitting

$$y = ax^2 + bx + c + \varepsilon$$



$$y = ax^{10} + bx^9 + cx^8 + \ldots \cancel{\varepsilon}$$

The risk of overfitting

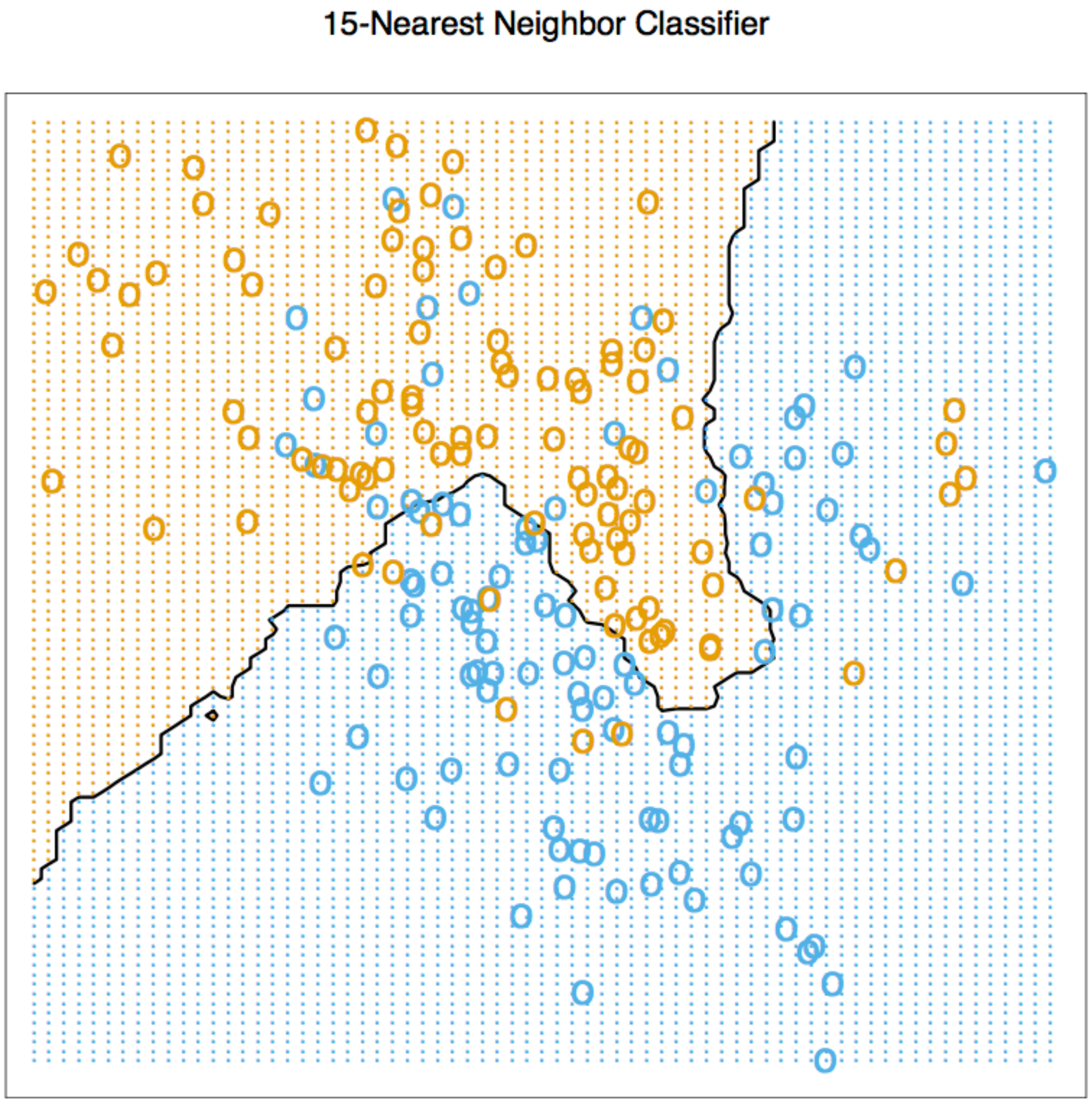
$$y = ax^2 + bx + c + \varepsilon$$

over-attending to noise

complex model

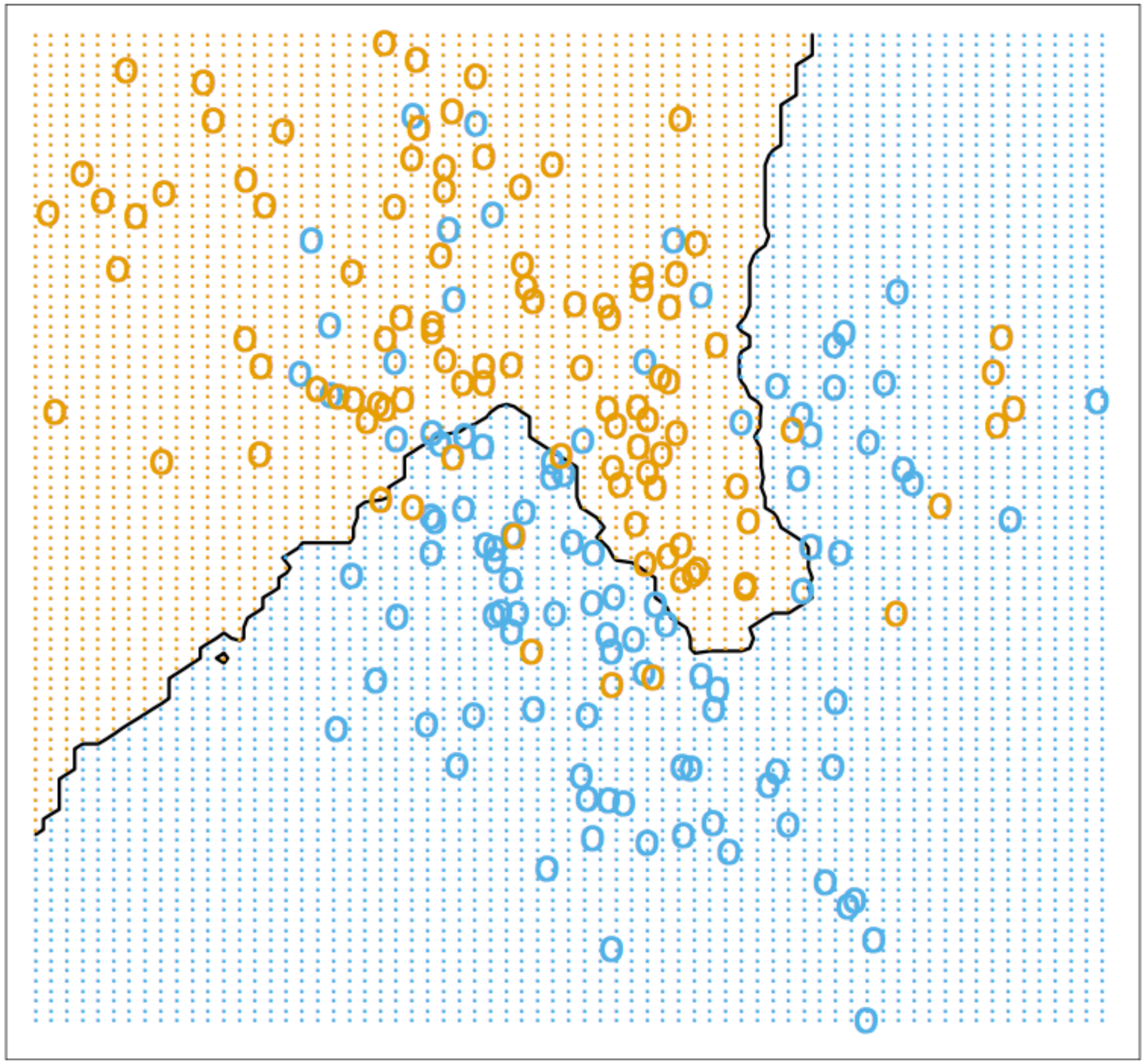
$$y = ax^{10} + bx^9 + cx^8 + \ldots \cancel{\varepsilon}$$

The risk of overfitting

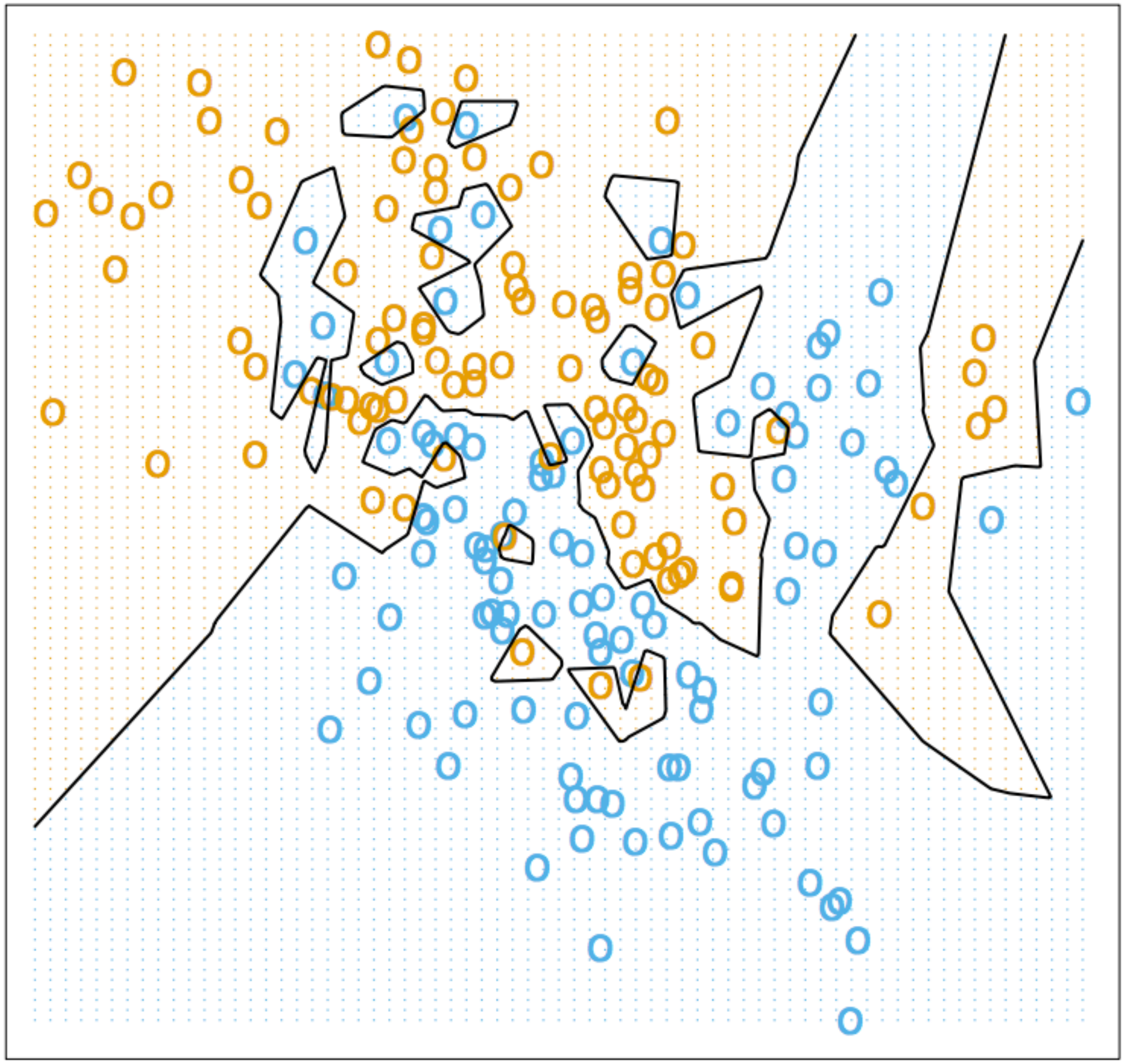


The risk of overfitting

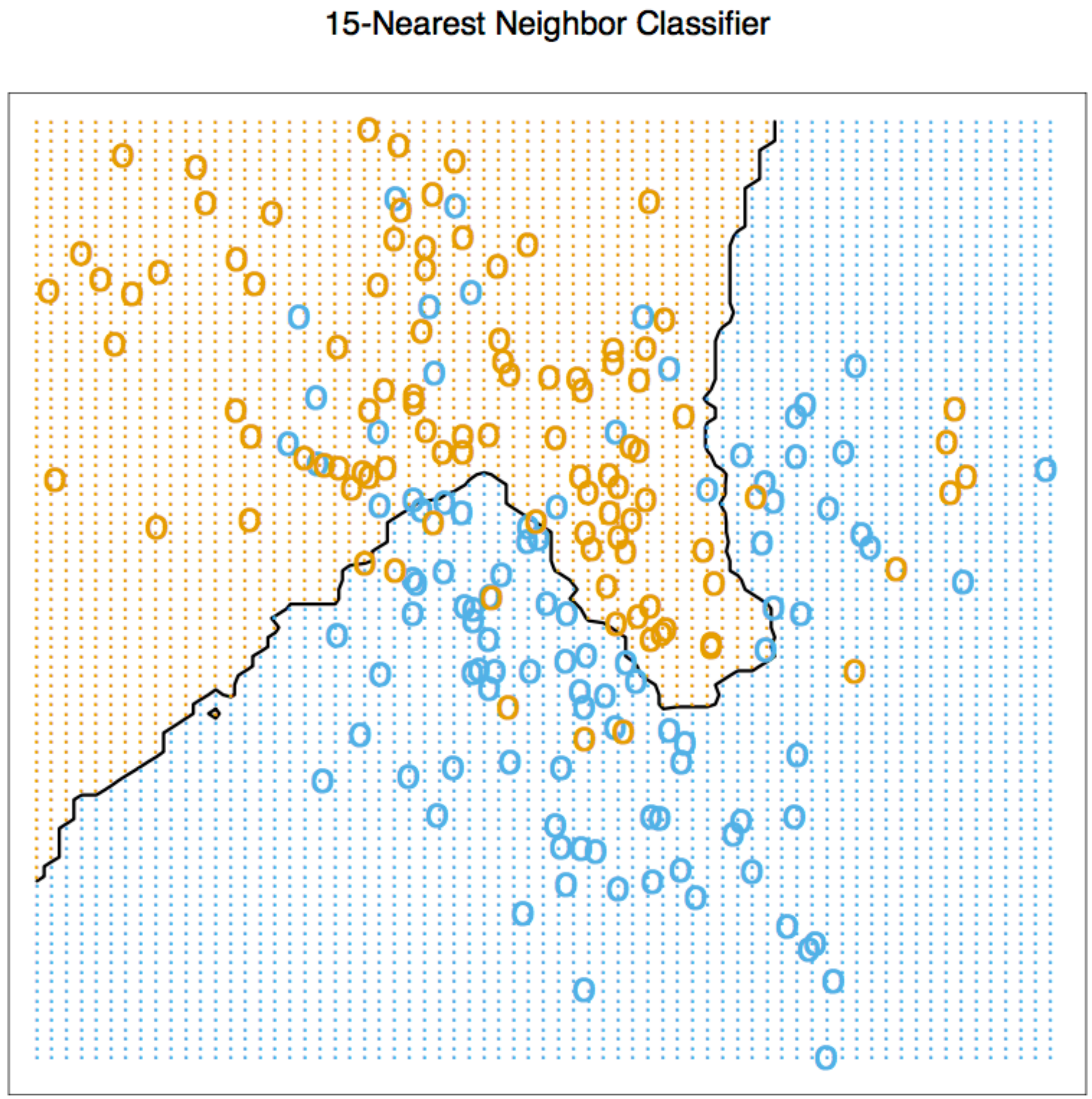
15-Nearest Neighbor Classifier

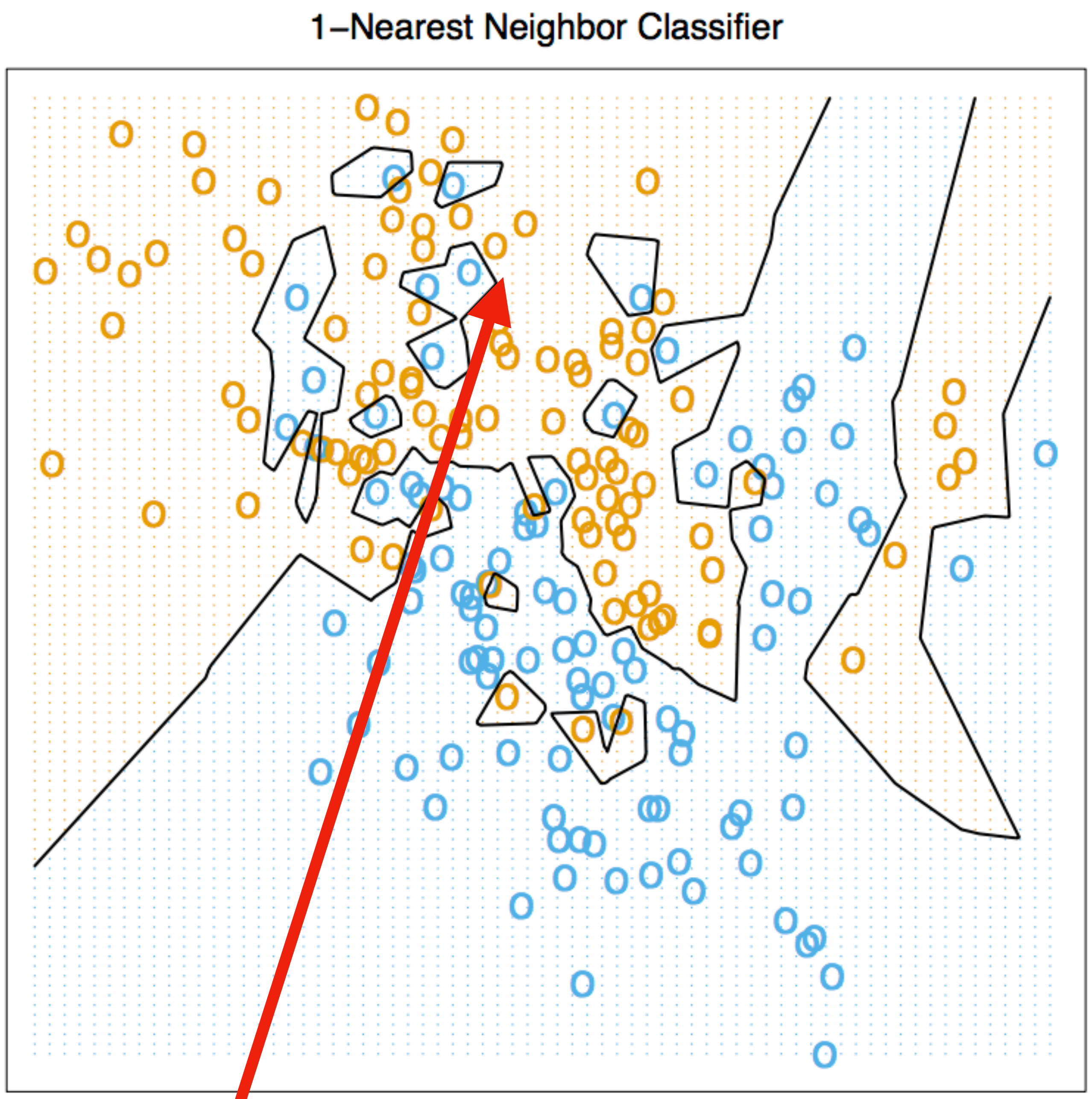


1–Nearest Neighbor Classifier



The risk of overfitting



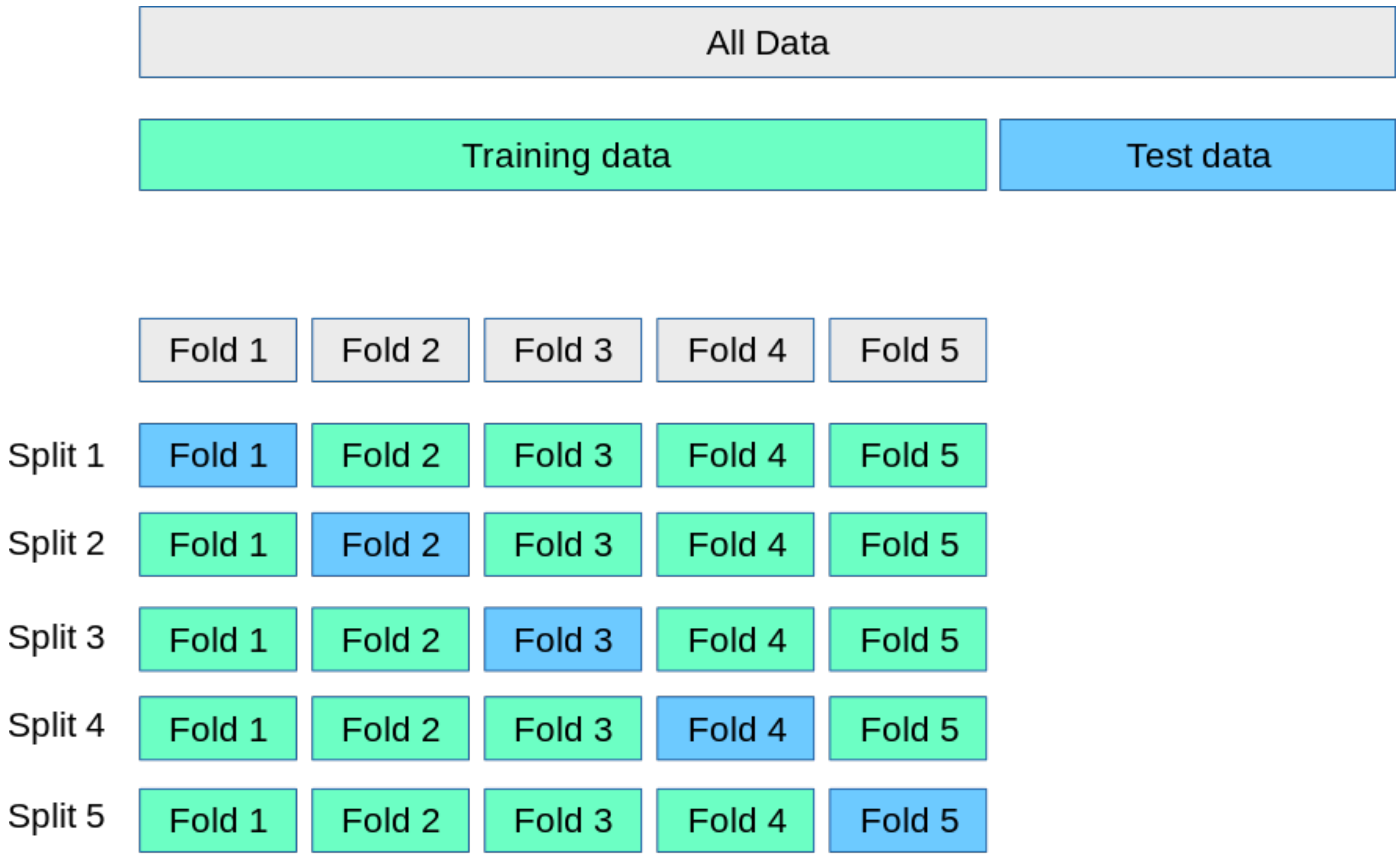


overly complex decision boundary

How well does a model generalize: Cross validation

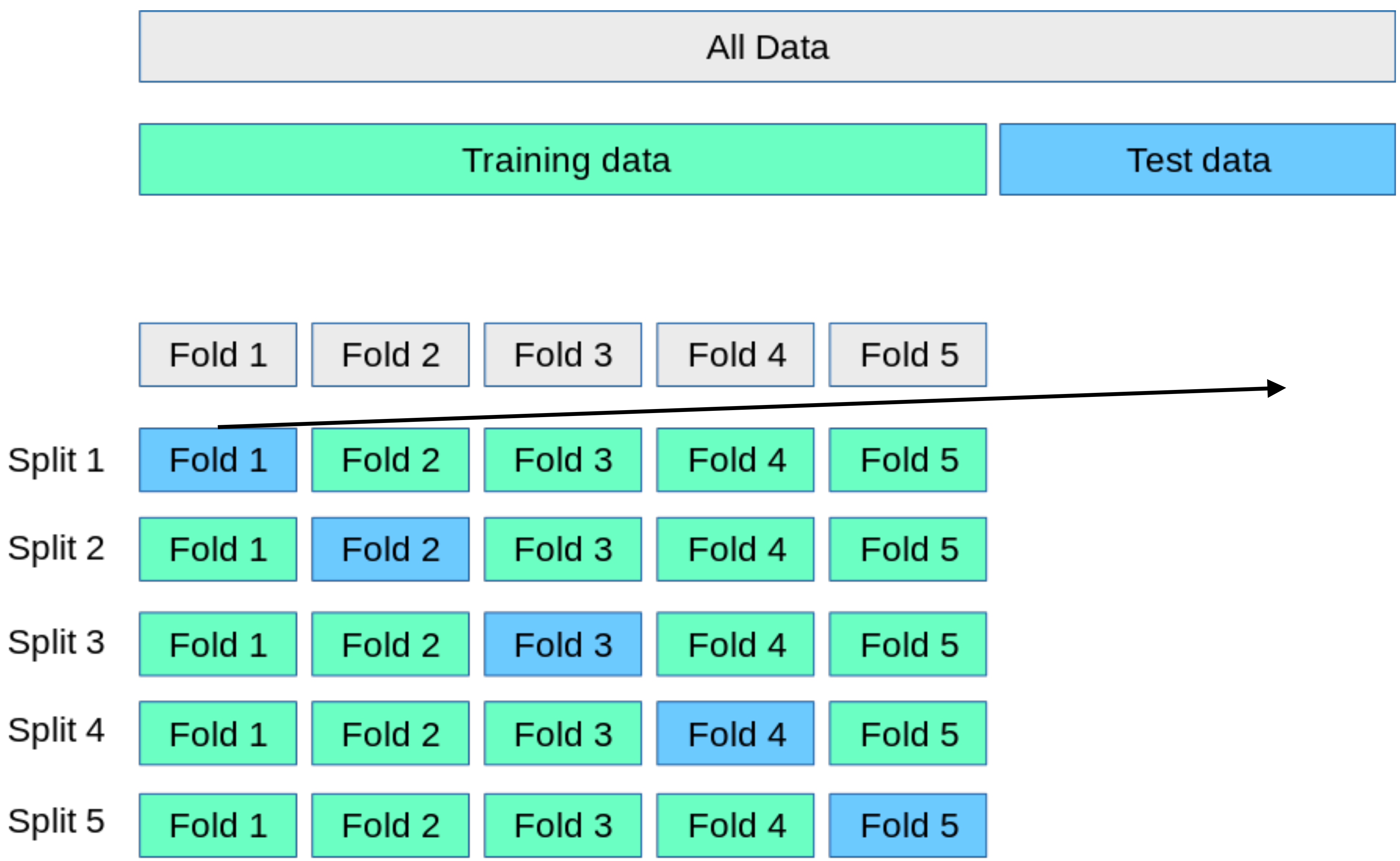
- A standard way of model evaluation in machine learning
    - Use some (larger) portion of data for training the model
    - Use the residual (smaller) portion for testing
    - Iterate the above procedures until every data point has been tested
    - The model is then assessed on the testing accuracy

Cross validation (5-fold)



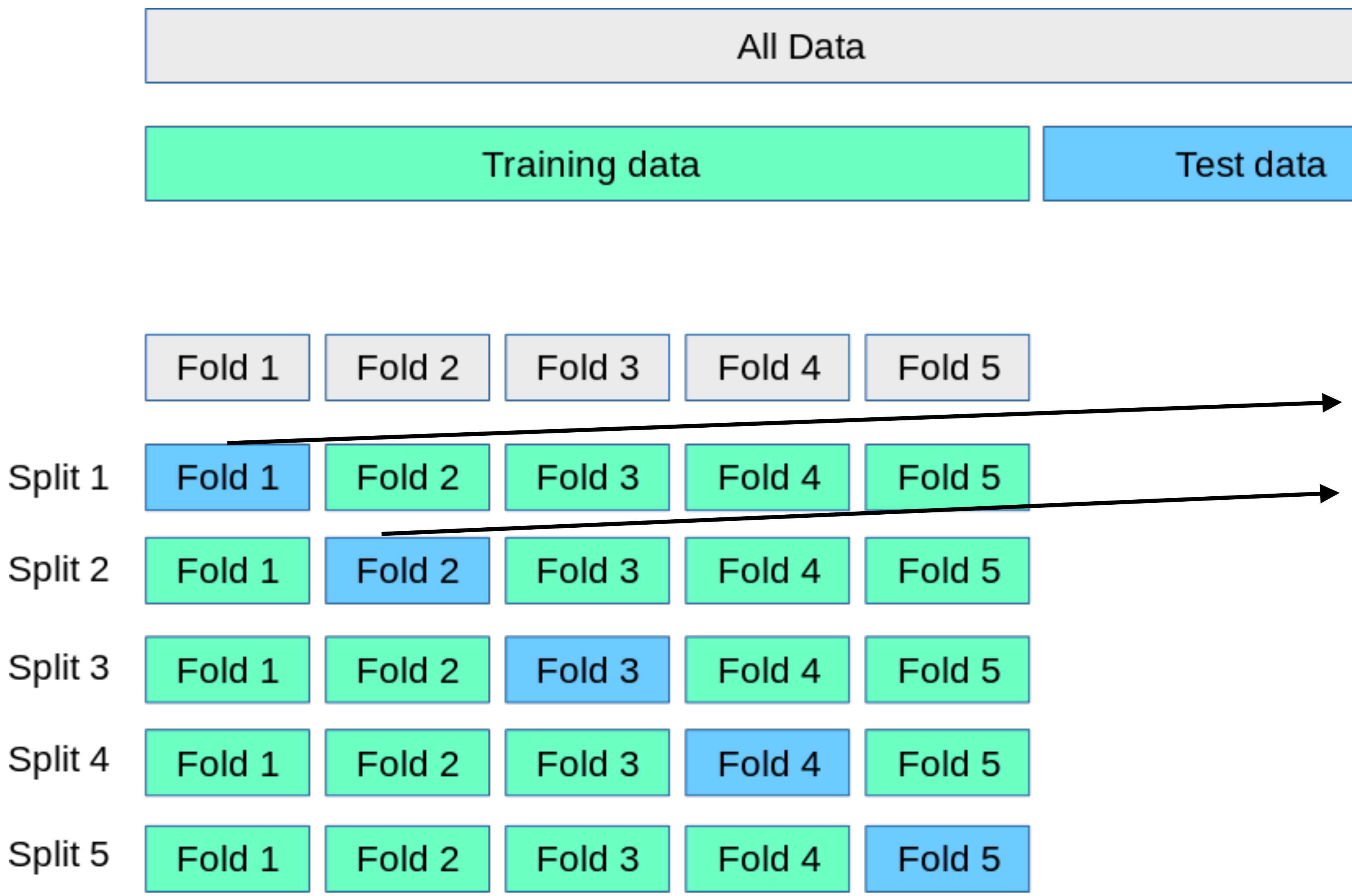
Adapted from scikit-learn

Cross validation (5-fold)



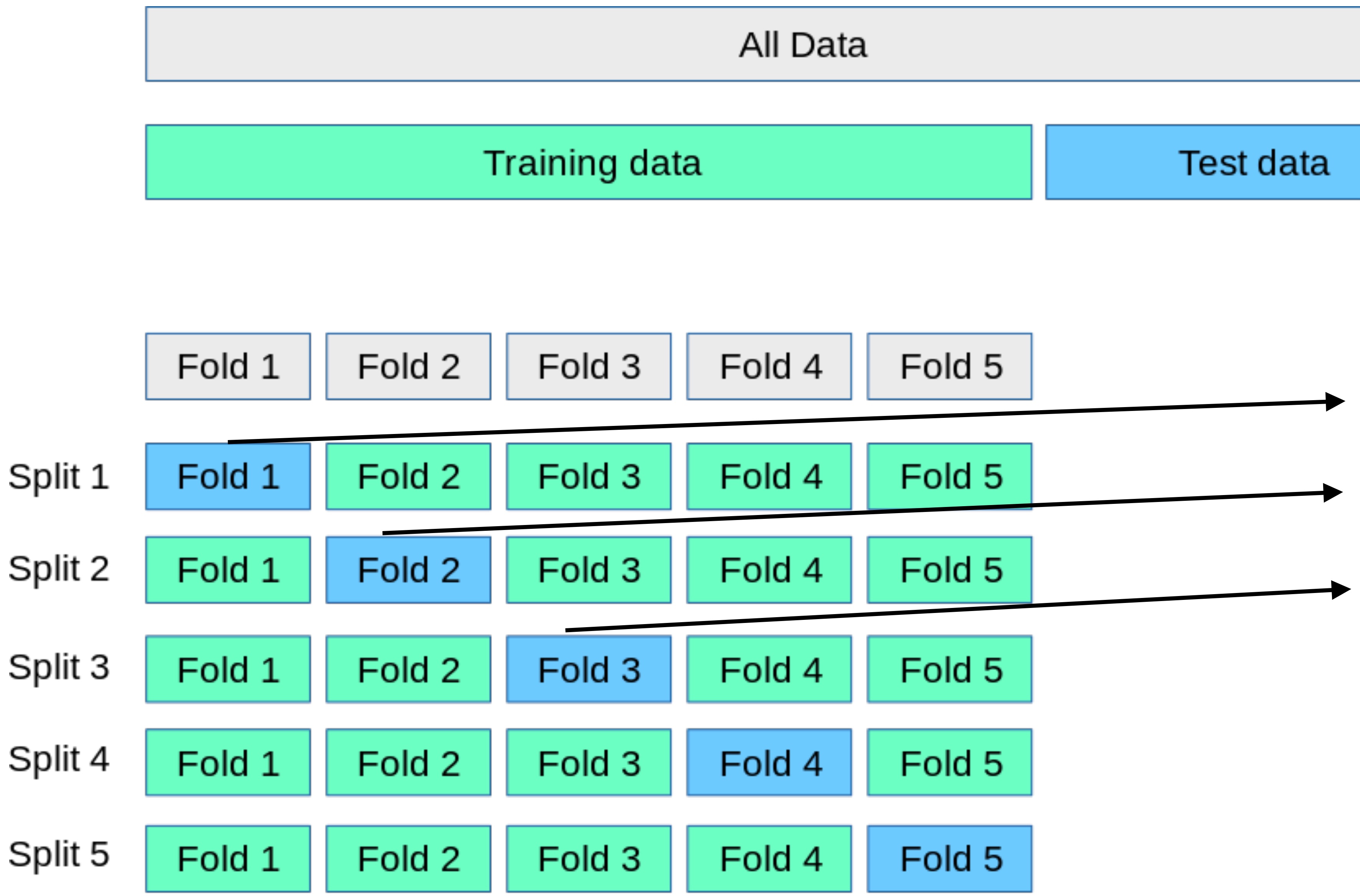
Adapted from scikit-learn

Cross validation (5-fold)



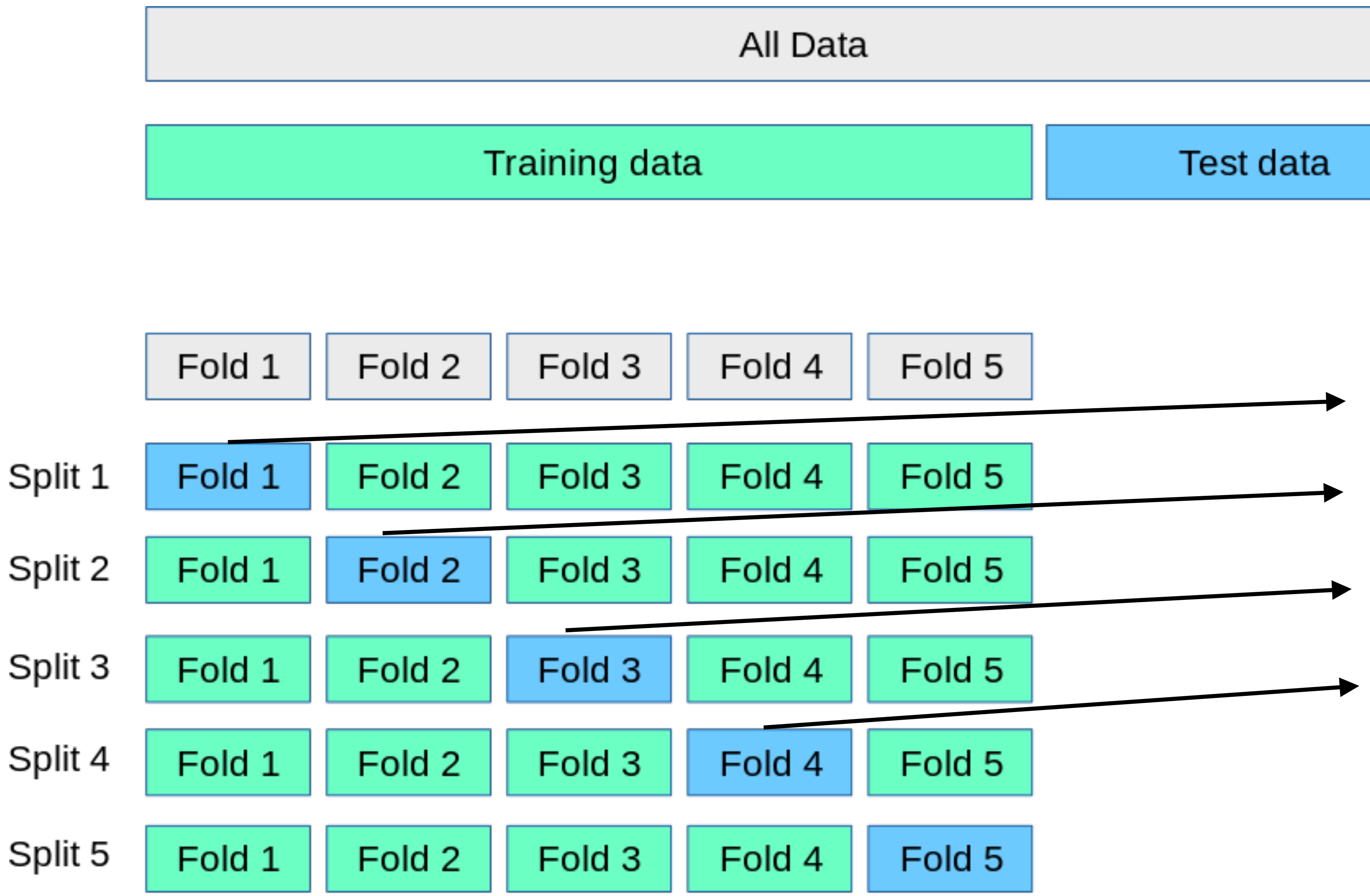
Adapted from scikit-learn

Cross validation (5-fold)



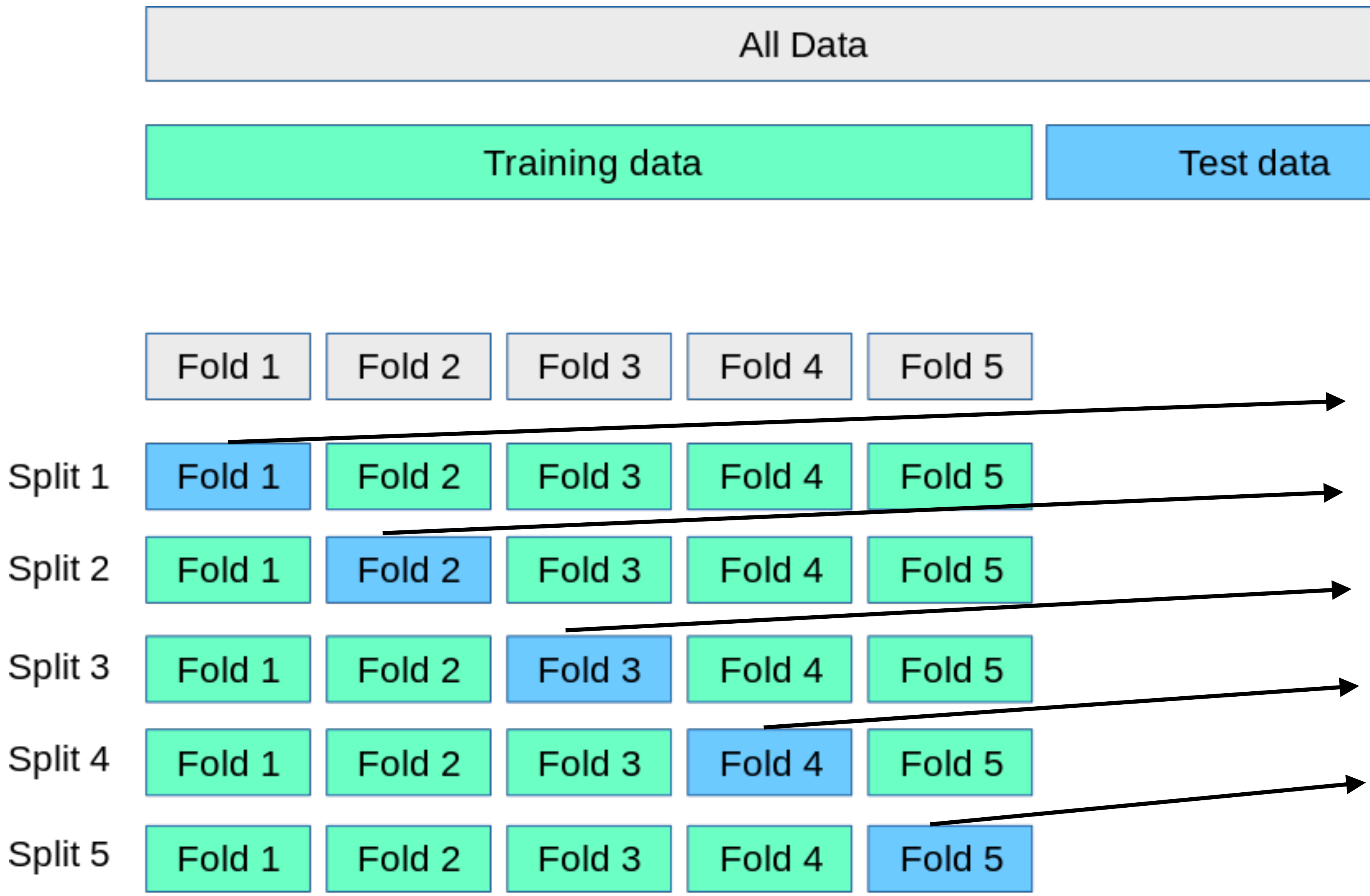
Adapted from scikit-learn

Cross validation (5-fold)



Adapted from scikit-learn

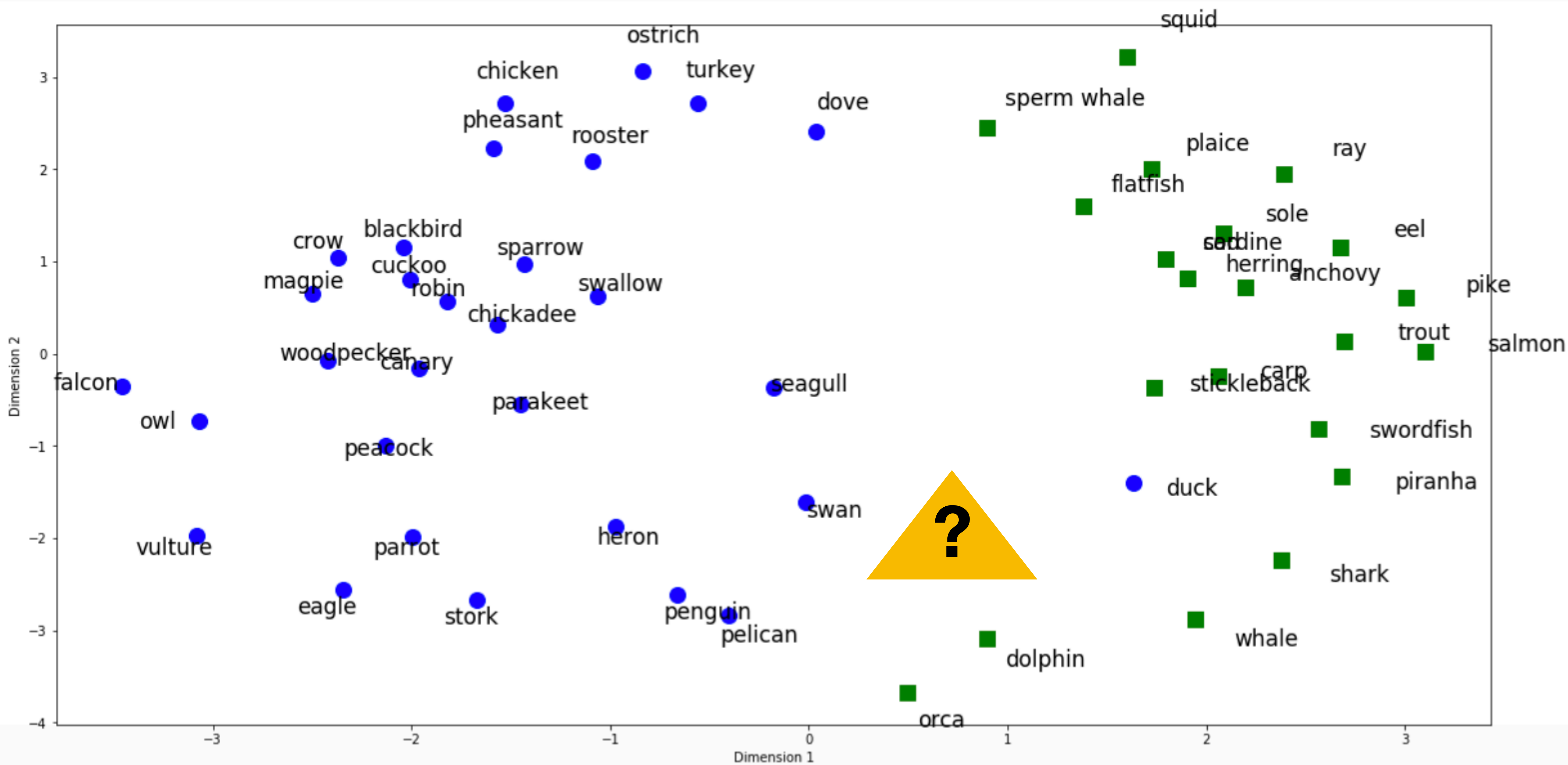
Cross validation (5-fold)



Adapted from scikit-learn

Leave-one-out cross validation: #folds = #data points

Category 1: Bird **Category 2: Fish**



Questions?

Question of the day

- How are complex categories formed over time?

Complex categories: A puzzle

- An example: Dyirbal (Australian) noun classifiers

I. *Bayi:* men, kangaroos, possums, bats, most snakes, most fishes, some birds, most insects, the moon, storms, rainbows, boomerangs, some spears, etc.

II. *Balan:* women, bandicoots, dogs, platypus, echidna, some snakes, some fishes, most birds, fireflies, scorpions, crickets, the hairy mary grub, anything connected with water or fire, sun and stars, shields, some spears, some trees, etc.

III. *Balam:* all edible fruit and the plants that bear them, tubers, ferns, honey, cigarettes, wine, cake

IV. *Bala:* parts of the body, meat, bees, wind, yamsticks, some spears, most trees, grass, mud, stones, noises and language, etc.

Dixon (1982); cf. Lakoff (1987)

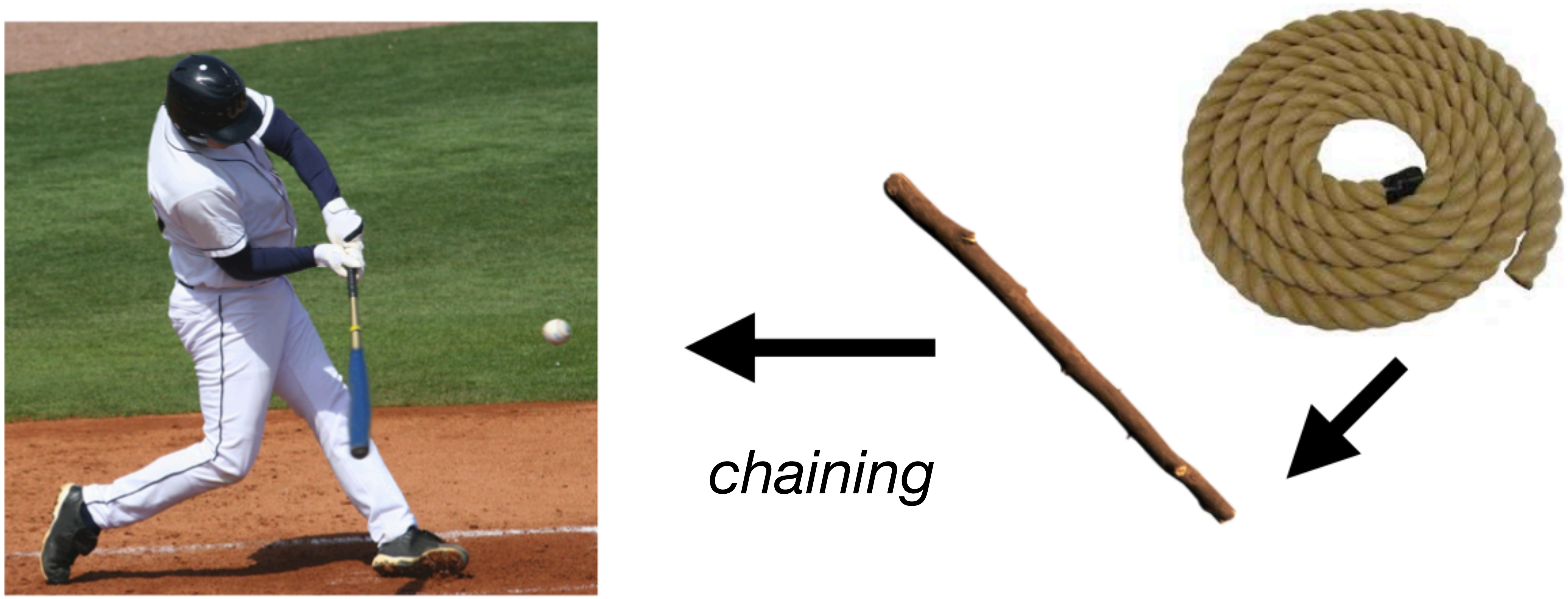
Complex categories: A puzzle

- Another example: Japanese classifier *hon*

Lakoff (1987)

Complex categories: A puzzle

- Another example: Japanese classifier *hon*



Lakoff (1987)

Complex categories: A puzzle

- Yet another example: English preposition *over*

  - Sam is walking *over* the hill.
  - Sam lives *over* the hill.
  - The wall fell *over*.
  - Sam turned the page *over*.
  - Sam turned *over*.
  - She spread the tablecloth *over* the table.
  - The guards were posted all *over* the hill.
  - The play is *over*.
  - Do it *over*, but don't *over*do it.
  - Look *over* my corrections, and don't *over*look any of them.
  - You made *over* a hundred errors.

Lakoff (1987)

Complex categories: A puzzle

- Yet another example: English preposition *over*

– Sam is walking *over* the hill.
– Sam lives *over* the hill.
– The wall fell *over*.
– Sam turned the page *over*.
– Sam turned *over*.
– She spread the tablecloth *over* the table.
– The guards were posted all *over* the hill.
– The play is *over*.
– Do it *over*, but don't *over*do it.
– Look *over* my corrections, and don't *over*look any of them.
– You made *over* a hundred errors.

Denoting Space

Time

Repetition

……

Lakoff (1987)

The problem of polysemy

Readings

Required reading:

– Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115(1), 39–57.

Optional readings

Optional readings:

- Reed, S. K. (1972). Pattern recognition and categorization. Cognitive Psychology, 3, 382–407.
- Medin, D. L., and Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85(3), 207–238.
- Medin, D. L., and Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, 35(1), 113–138.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological review*, 98(3), 409–429.
- Goldstone, R. L., and Kersten, A. (2003). Concepts and categorization, in Healy, A.F. and Proctor, R.W. (eds), *Comprehensive handbook of psychology, Volume 4: Experimental psychology*. Wiley, 599–621.

Recommended book:

- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago press.

Lab 5: Categorization