CS461 – RECITATION 03

MACHINE LEARNING PRINCIPLES

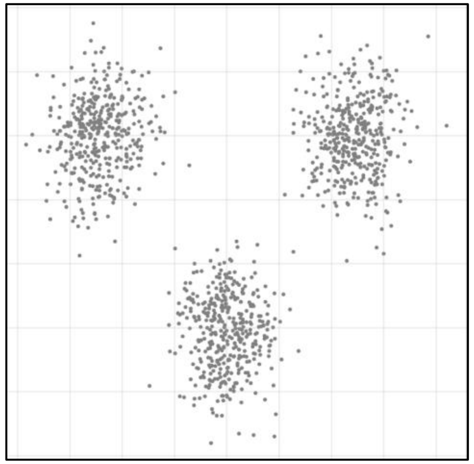
Daize Dong

2025-09-30

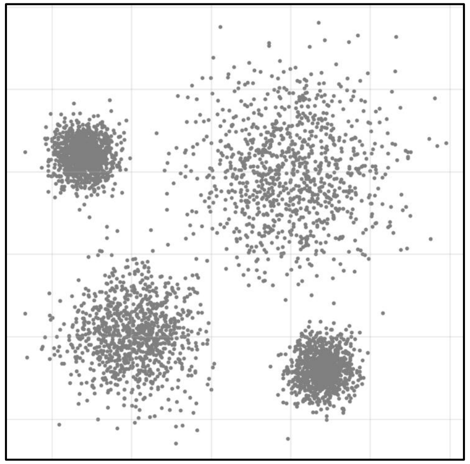
TODAY’S CONTENT

- K-means
- Quiz 01

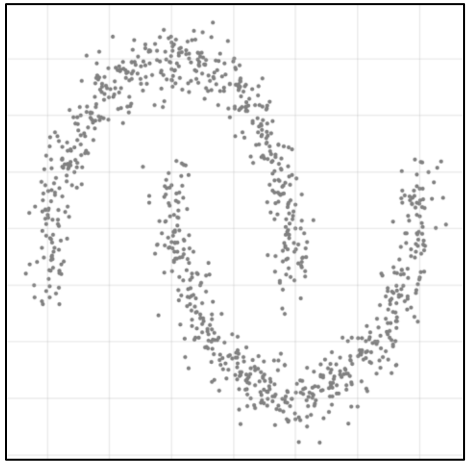
K-MEANS



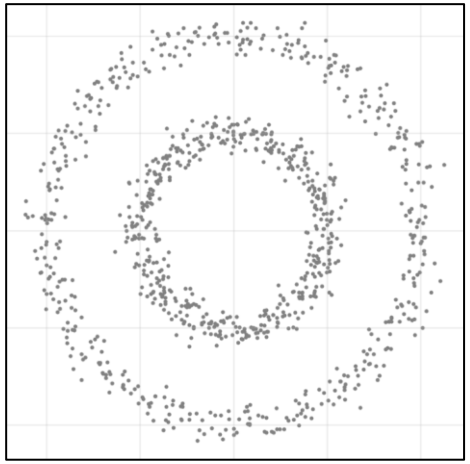
Blob



Gaussian

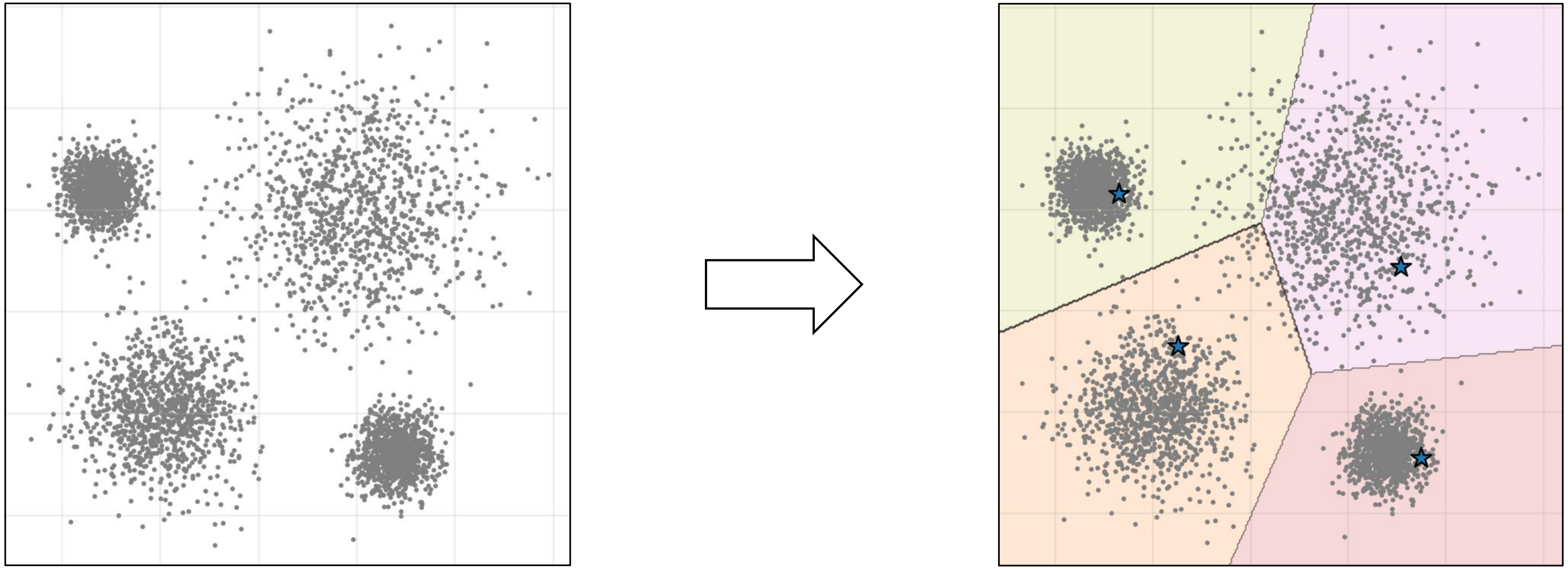


Moon



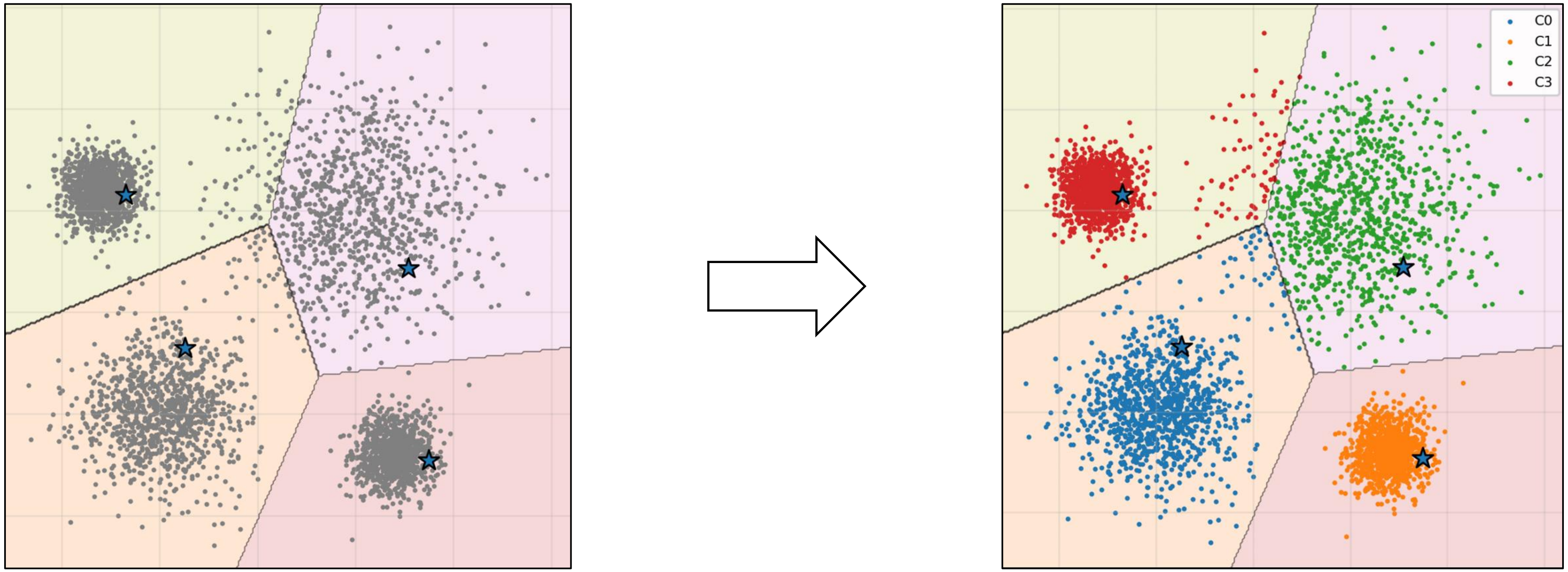
Circle

STEP 1:
INITIALIZE CENTROIDS



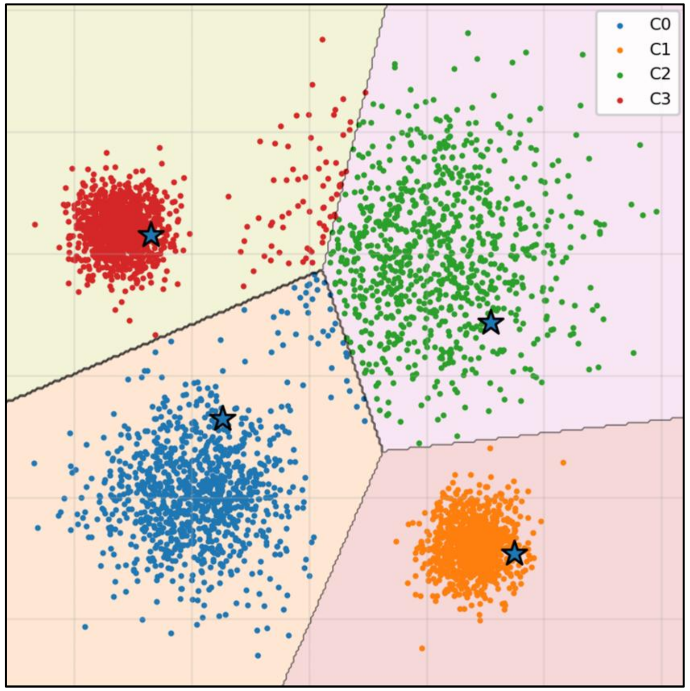
Choose K value and randomly pick K points

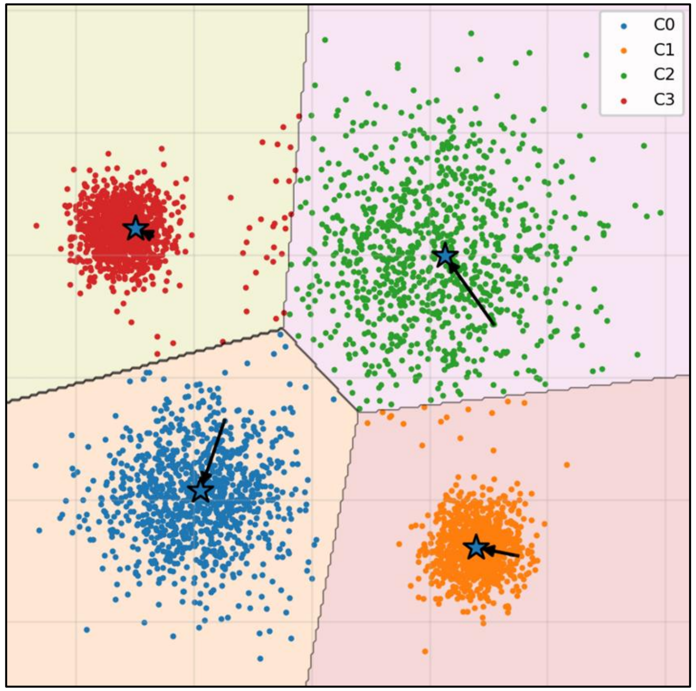
STEP 2:
ASSIGN NEAREST CENTROIDS



Calculate the distance to each centroid for all data

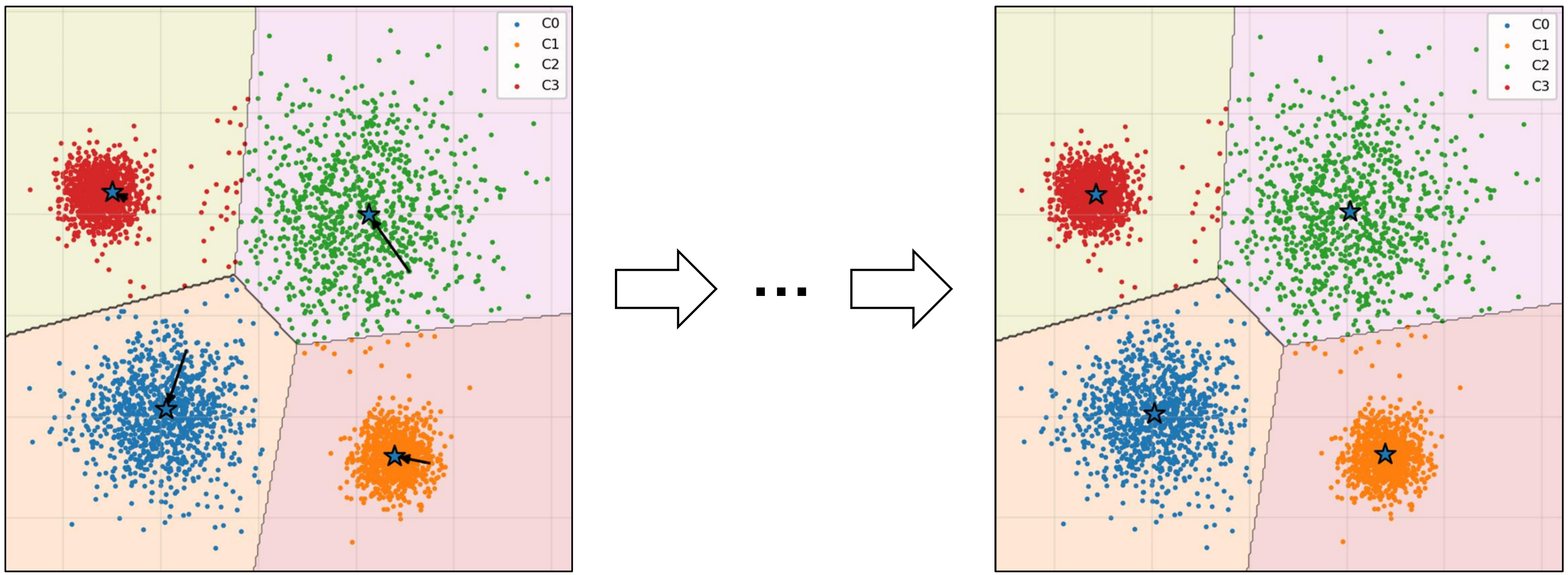
STEP 3:
GET NEW CENTROIDS





New centroid: mean value of all data within the cluster

STEP 4:
REPEAT UNTIL STABLE

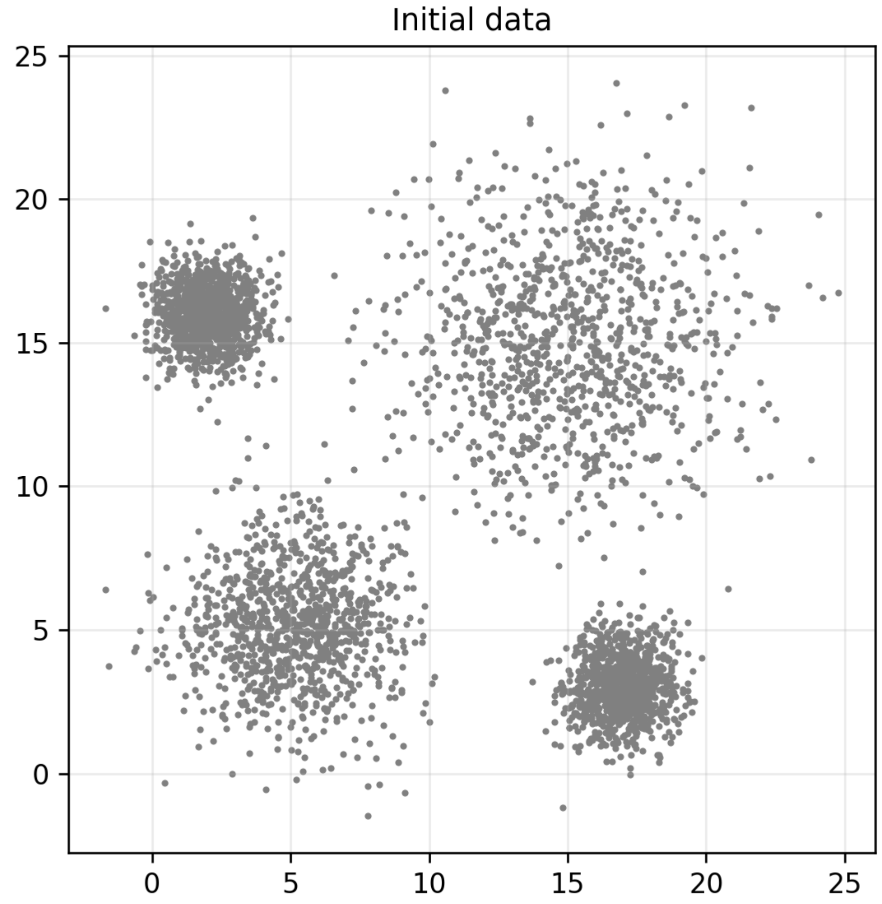


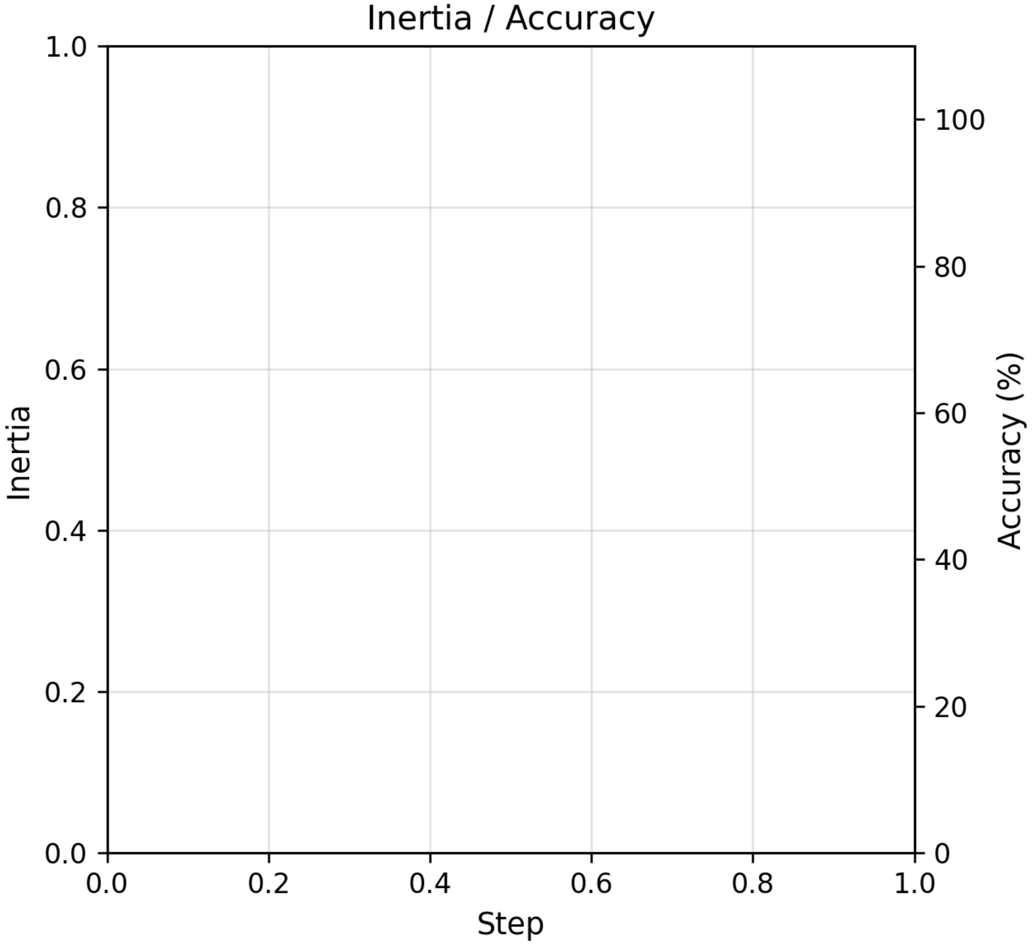
Stable: No/Subtle change on each cluster

GAUSSIAN DATA

gaussians — K-Means (k=4) — initial data (unclustered)

GIF (best view in .pptx)

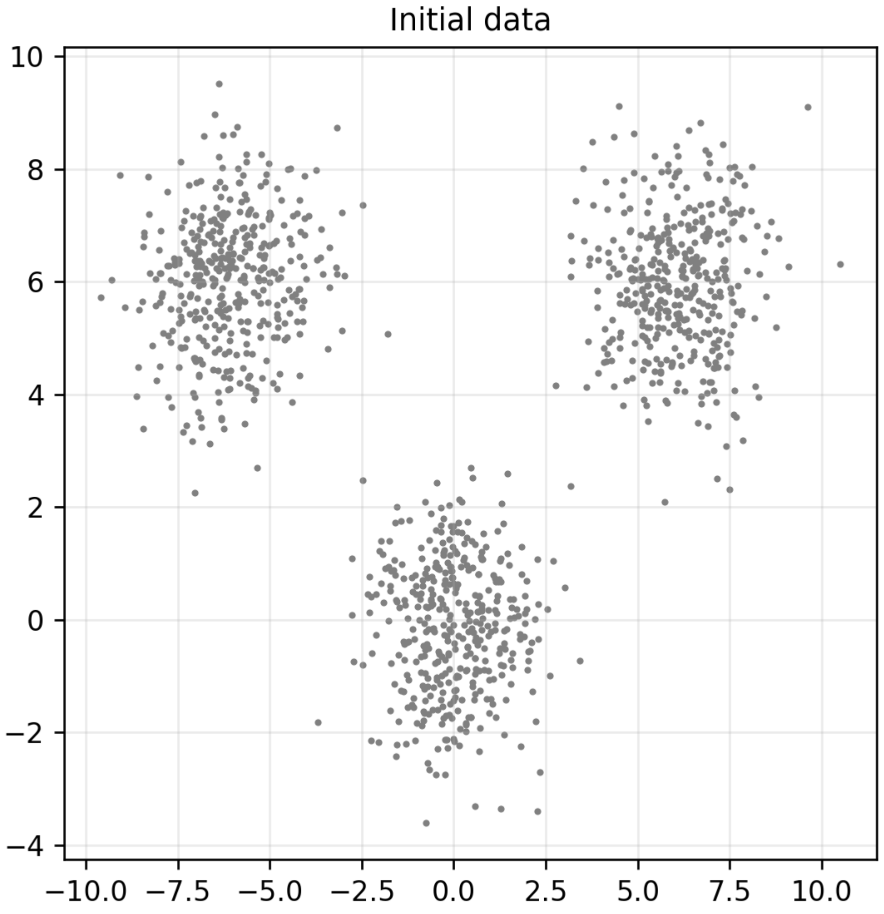


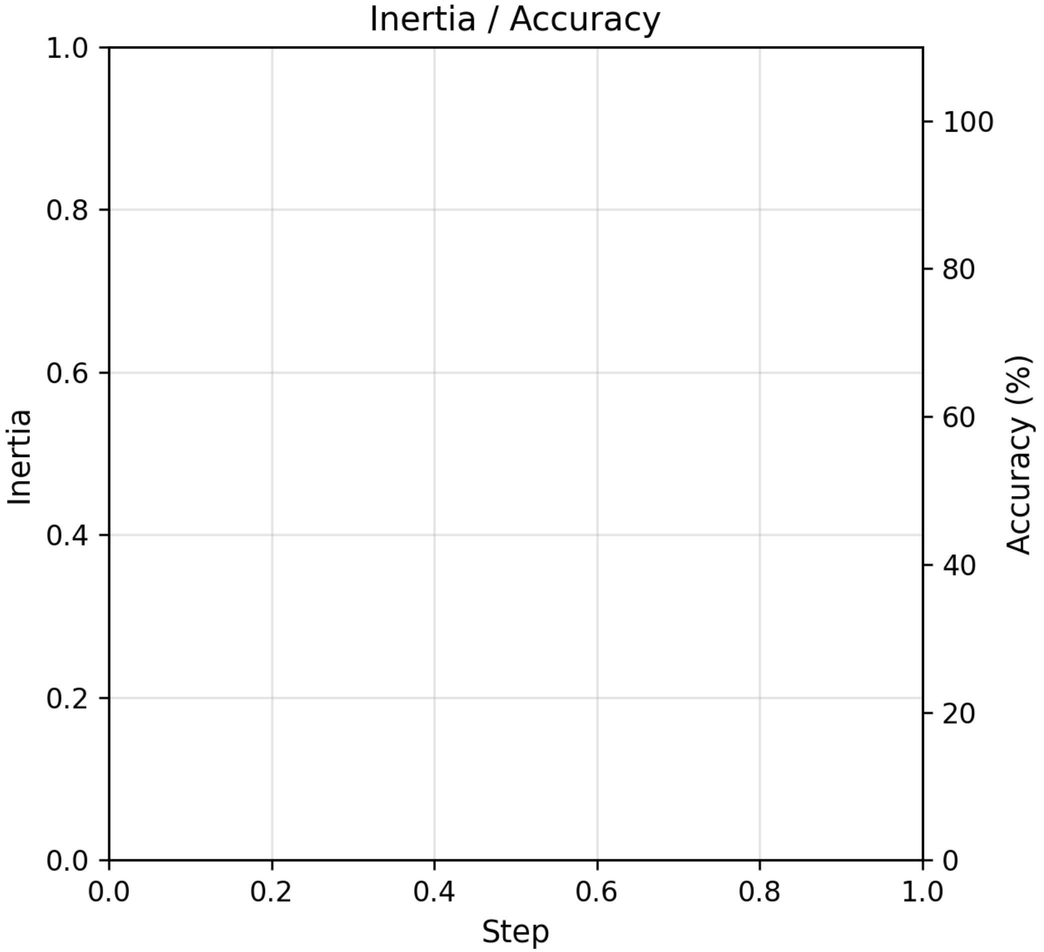


BLOB DATA

blobs — K-Means (k=3) — initial data (unclustered)

GIF (best view in .pptx)

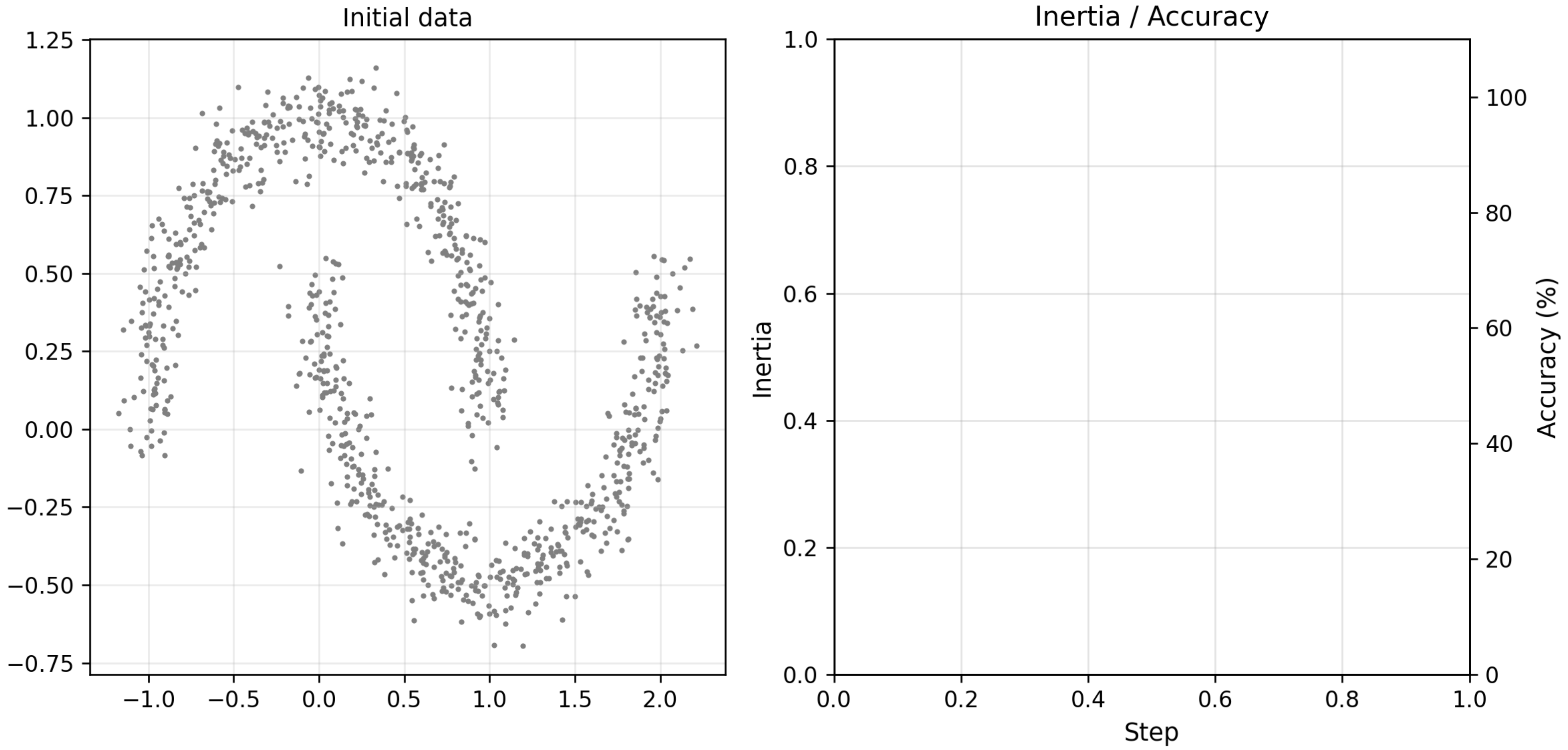




MOON DATA

moons — K-Means (k=2) — initial data (unclustered)

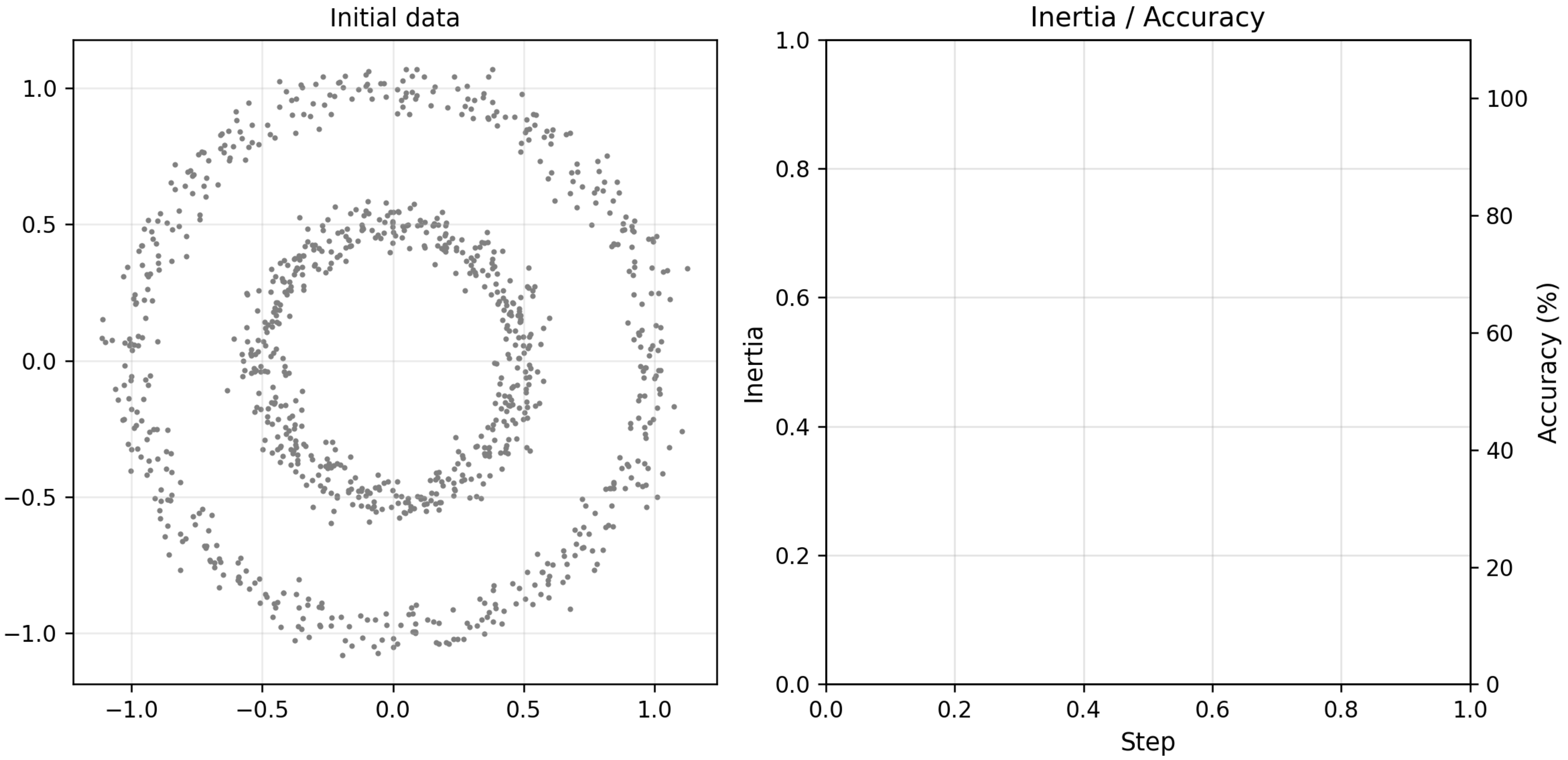
GIF (best view in .pptx)



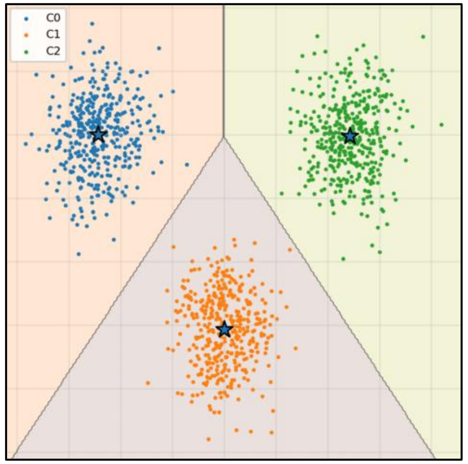
CIRCLE DATA

circles — K-Means (k=2) — initial data (unclustered)

GIF (best view in .pptx)

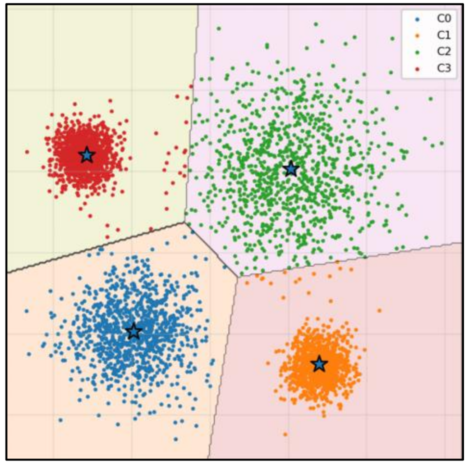


SUMMARY



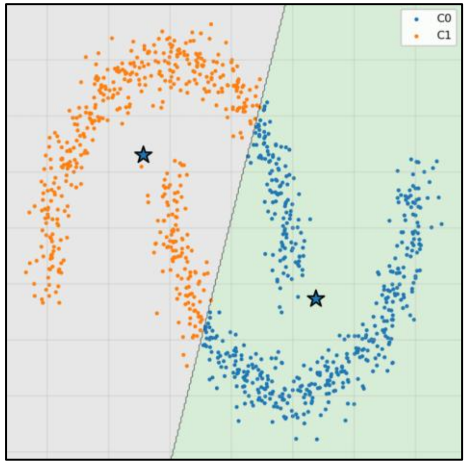
Blob

100% ✅



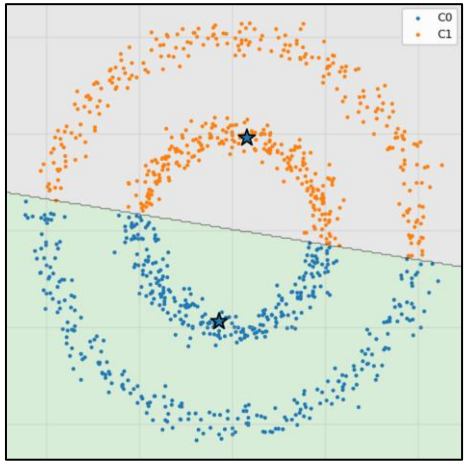
Gaussian

99% ✅



Moon

75% 💦



Circle

50% 🤡

QUIZ 01

A	B	C	D
6	0	14	0

1) Which of the following statements are true?

a. Unbiased estimators have the lowest error of any estimators.

b. The sin function is not polynomial so bias cannot be measured.

c. 0 is an unbiased estimator of sin(x)

d. The sin function exists from $-\infty$ to ∞ so bias cannot be measured.

Mean Squared Error (MSE) of an estimator $\hat{\theta}$:

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta}-\theta)^2\right].$$

$$= \underbrace{(\mathbb{E}[\hat{\theta}]-\theta)^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}\left[(\hat{\theta}-\mathbb{E}[\hat{\theta}])^2\right]}_{\text{Variance}} + \underbrace{\sigma^2_{\text{irreducible}}}_{\text{Noise floor}}$$

Biased, regularized estimators can have lower Mean Square Error (MSE).

QUIZ 01

A	B	C	D
3	10	2	5

2) As the amount of noise increases, which of the following is false?

a. More data is required to estimate the signal.

b. More computationally expensive machine learning techniques are required.

c. The quality of the signal estimate worsens.

d. If the noise is i.i.d., then the mean of the samples is still the best estimate of the signal's mean.

$$y_i = \mu + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i] = 0, \ \mathrm{Var}(\varepsilon_i) = \sigma^2$$

$$\mathbb{E}[\mu] = \mu$$

More compute ≠ Lower variance
Sometimes it makes the fitting worse

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \mathbb{E}[\bar{y}] = \mu, \quad \mathrm{Var}(\bar{y}) = \frac{\sigma^2}{n}$$

QUIZ 01

A	B	C	D
10	7	3	0

3) Which of these sentences best describes the bias variance trade-off?

a. The person modelling should choose a bias-variance trade-off based upon knowledge of the dataset.

b. A model should capture the variance in a dataset by increasing its complexity and reducing bias to a presumed distribution.

c. A model should be simplified to bias it to avoid matching the variance of a dataset.

d. The bias-variance trade-off is inherent to a dataset and is not controlled by the model.

Not all dataset are the same so we need to choose by case
Final goal: lower the **error**

QUIZ 01

A	B	C	D
0	1	13	6

4) When we modify the normal equation like this, β = $(XTX + \lambda I)-1XTy$, the λ

a. is a meaningless constant.

b. is only there to stabilize the matrix inversion.

c. shapes the parameters, biasing the model to better match noise of a guassian distribution.

d. is a heuristic that was invented by Laplace to make Guass' least squares regression work on non-Guassian data.

Basic concept in ridge regression

QUIZ 01

A	B	C	D
2	9	7	2

5) Which of the following is false about decision boundaries?

a. For samples on the decision boundary, we expect the probability of belonging to a class to be near 0.5.

b. Logistic regression will create decision boundaries based upon sample statistics, not prediction error rates.

c. The perceptron model will create decision boundaries based upon sample statistics, not prediction error rates.

d. The sigmoid function converts a decision boundary to a probability estimate.

- Logistic regression fits parameters by maximizing likelihood.
  Boundary comes from fitted probabilities ($p_0 = 0.4, p_1 = 0.6$), not directly from minimizing classification errors.
- The perceptron model only looks at the prediction error (-1 or 1), and not at sample statistics.

QUIZ 01

6) Build a decision tree from the following data, predicting the class from x and y:

class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6

Use the Gini impurity, $1 - \sum_{c=1}^{C} \hat{p}_c^2$. At each node on the tree, write either the pivot value or, for the leaf nodes, the class. If multiple pivots have the same Gini impurity, you may choose any one of them arbitrarily. Splits are $<$ and $\geq$. You do not need to calculate every value; try plotting them if you are confused.

QUIZ 01

```
def gini_impurity(classlist):
    ......
    # Calculate the geni impurity for a single set
    classes, counts = np.unique(classlist, return_counts=True)
    p_hats = counts / len(classlist)
    gini = 1.0 - np.sum(p_hats ** 2)
    return np.sqrt(gini)

def get_best_impurity(column, classlist, comparison_left, comparison_right):
    ......
    # The weighted gini impurity of both sides
    left_impurity = gini_impurity([classlist[index] for index in left_indices])
    right_impurity = gini_impurity([classlist[index] for index in right_indices])
    weighted_impurity = (len(left_indices) / len(classlist)) * left_impurity + \
                        (len(right_indices) / len(classlist)) * right_impurity
    ......
```

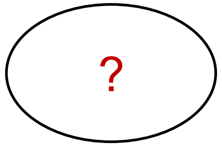
Gini Impurity:

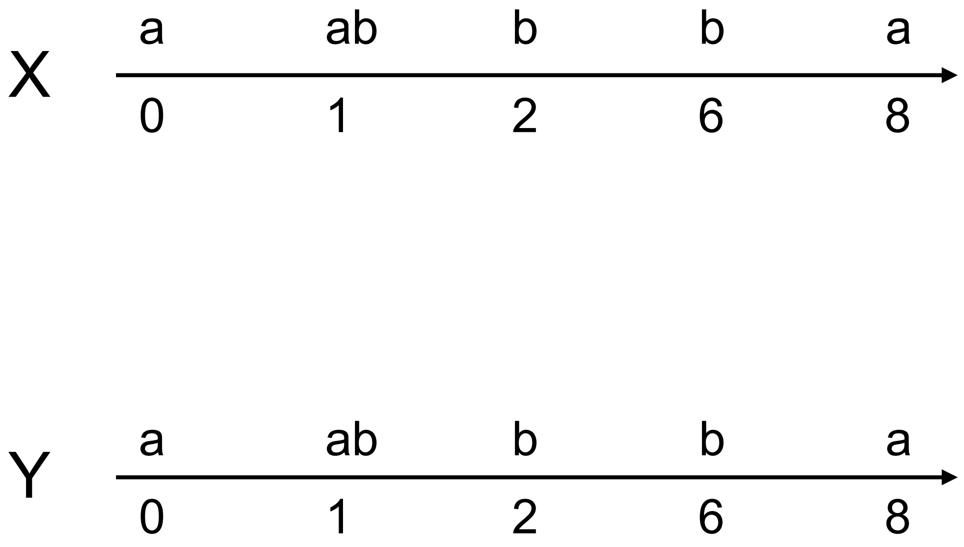
$$1 - \sum_{c=1}^{C} \hat{p}_c^2.$$

This decides the split point for attributes!

QUIZ 01

class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6

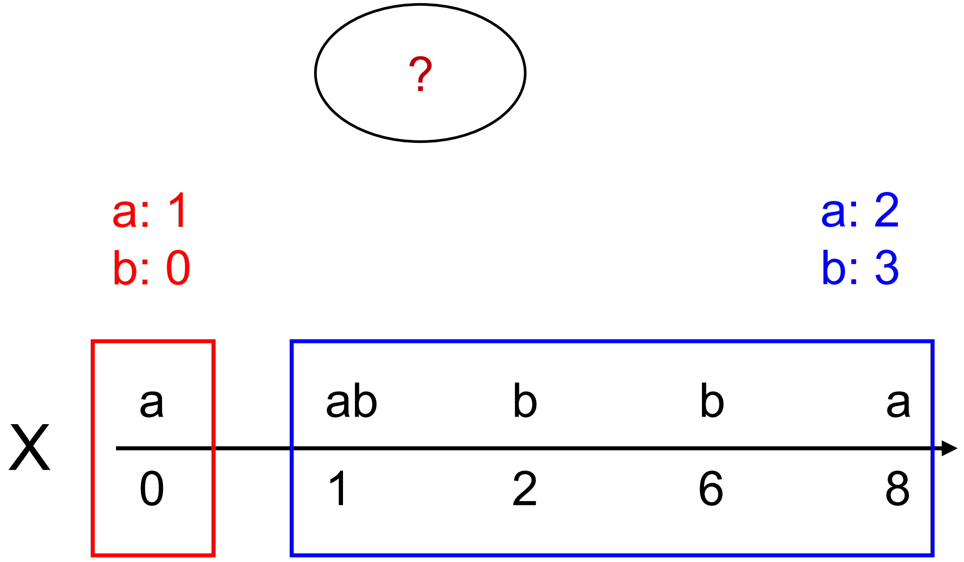




Initialize points:

QUIZ 01

class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6



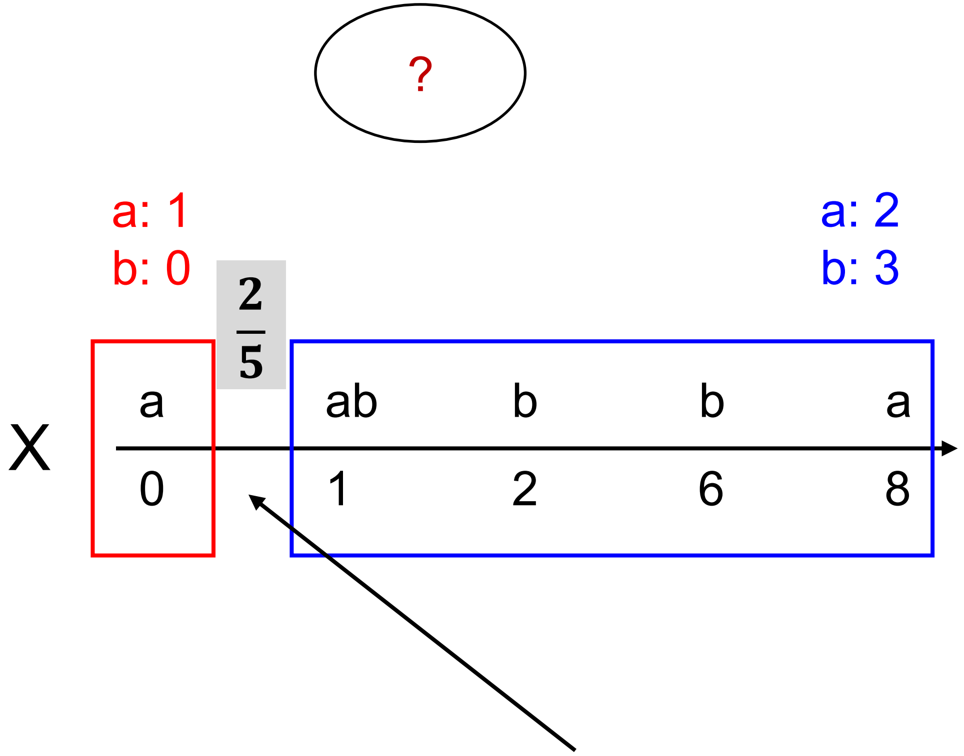
$$1-\sum_{c=1}^{C}\hat{p}_c^2.$$

Single gini impurity: $1-\left(\frac{1}{1+0}\right)^2=\mathbf{0}$ $\quad 1-\left(\frac{2}{2+3}\right)^2-\left(\frac{3}{2+3}\right)^2=\mathbf{\frac{12}{25}}$

Weighted gini impurity: $\frac{1}{6}\times\mathbf{0}+\frac{5}{6}\times\mathbf{\frac{12}{25}}=\mathbf{\frac{2}{5}}$

QUIZ 01

class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6

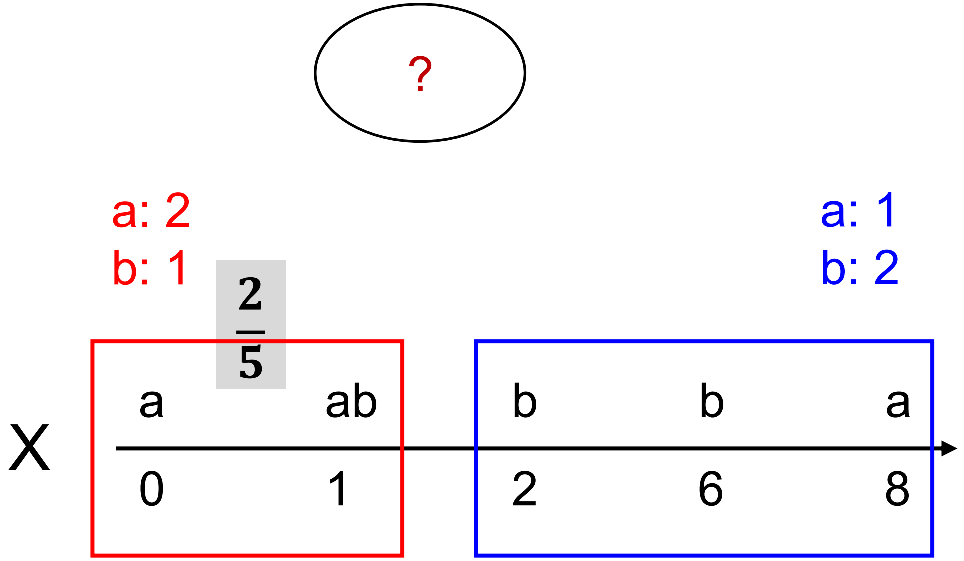


Weighted gini impurity:

$$\frac{1}{6} \times \mathbf{0} + \frac{5}{6} \times \frac{\mathbf{12}}{\mathbf{25}} = \frac{\mathbf{2}}{\mathbf{5}}$$

QUIZ 01

class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6



$$1-\sum_{c=1}^{C}\hat{p}_c^2.$$

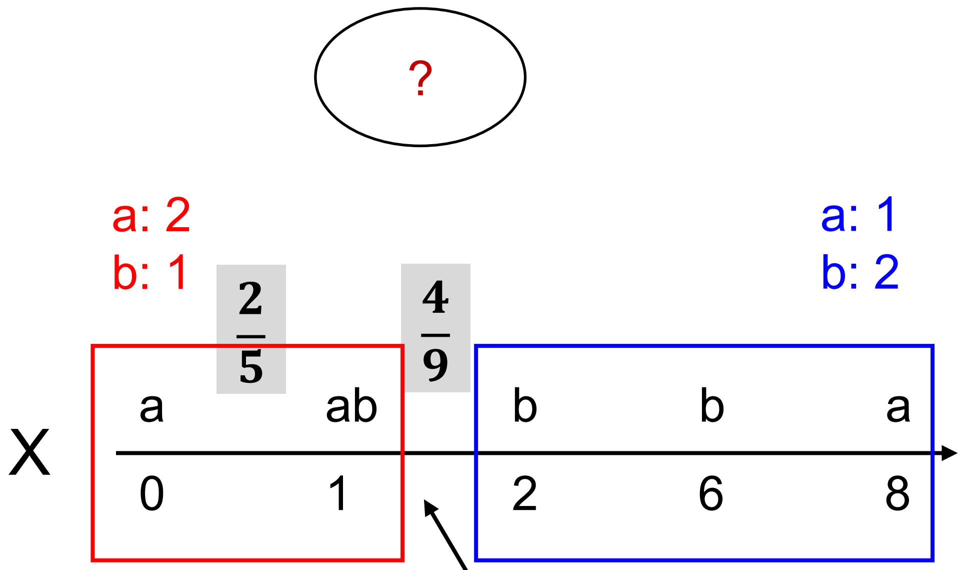
Continue with others:

$$1-\left(\frac{2}{1+2}\right)^2-\left(\frac{1}{1+2}\right)^2=\mathbf{\frac{4}{9}} \qquad 1-\left(\frac{1}{1+2}\right)^2-\left(\frac{2}{1+2}\right)^2=\mathbf{\frac{4}{9}}$$

$$\frac{3}{6}\times\mathbf{\frac{4}{9}}+\frac{3}{6}\times\mathbf{\frac{4}{9}}=\mathbf{\frac{4}{9}}$$

QUIZ 01

class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6

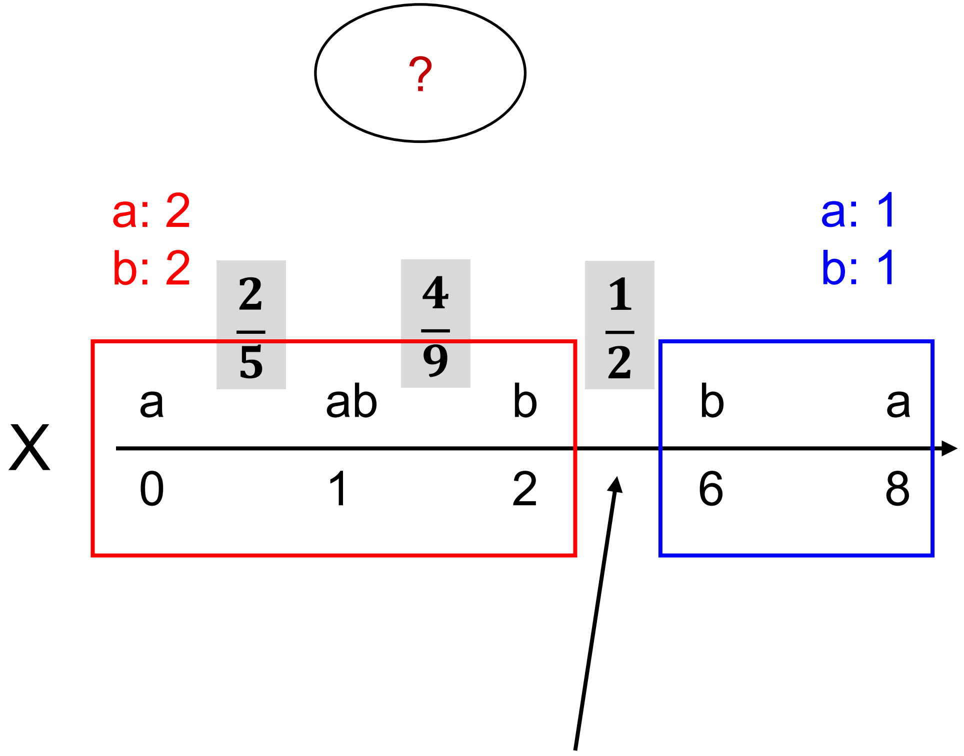


$$\frac{3}{6}\times\frac{4}{9}+\frac{3}{6}\times\frac{4}{9}=\frac{4}{9}$$

Continue with others:

QUIZ 01

class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6

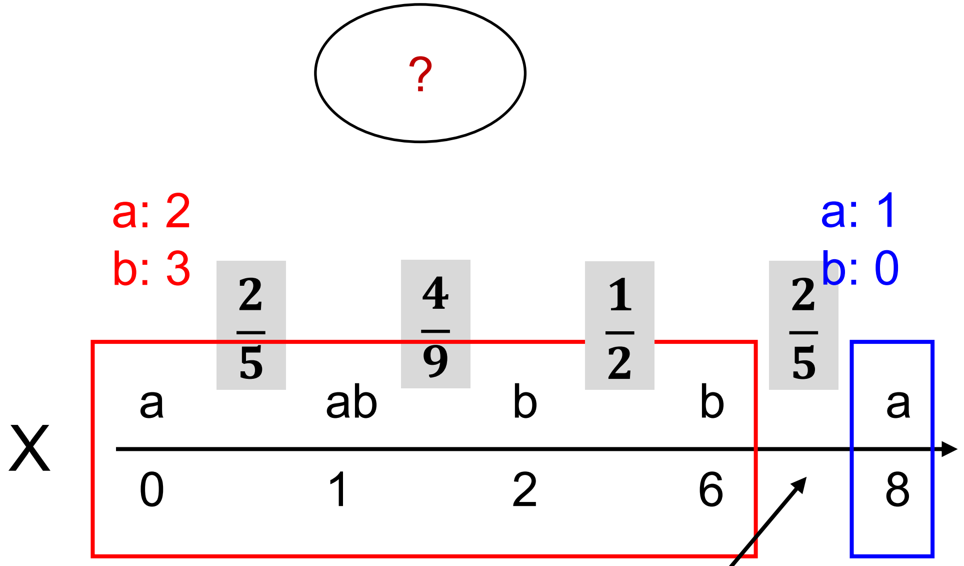


$$\frac{4}{6}\times\frac{1}{2}+\frac{2}{6}\times\frac{1}{2}=\frac{1}{2}$$

Keep going on:

QUIZ 01

class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6

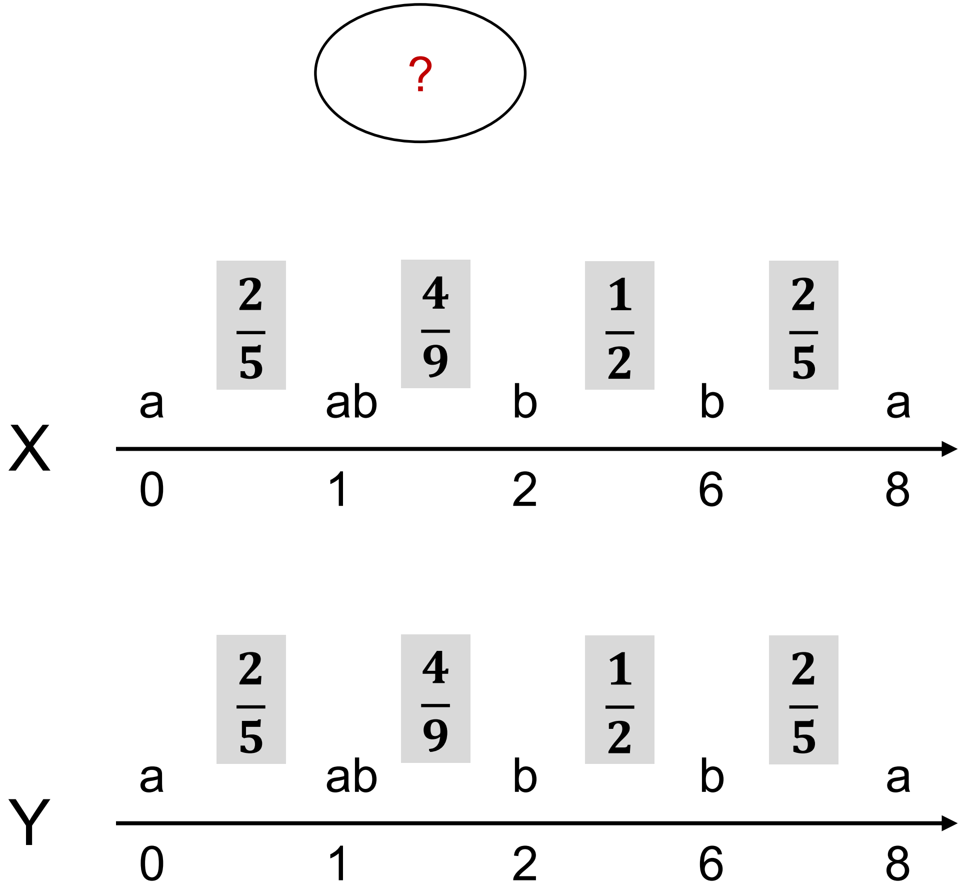


Keep going on:

$$\frac{5}{6} \times \frac{\mathbf{12}}{\mathbf{25}} + \frac{1}{6} \times \mathbf{0} = \frac{\mathbf{2}}{\mathbf{5}}$$

QUIZ 01

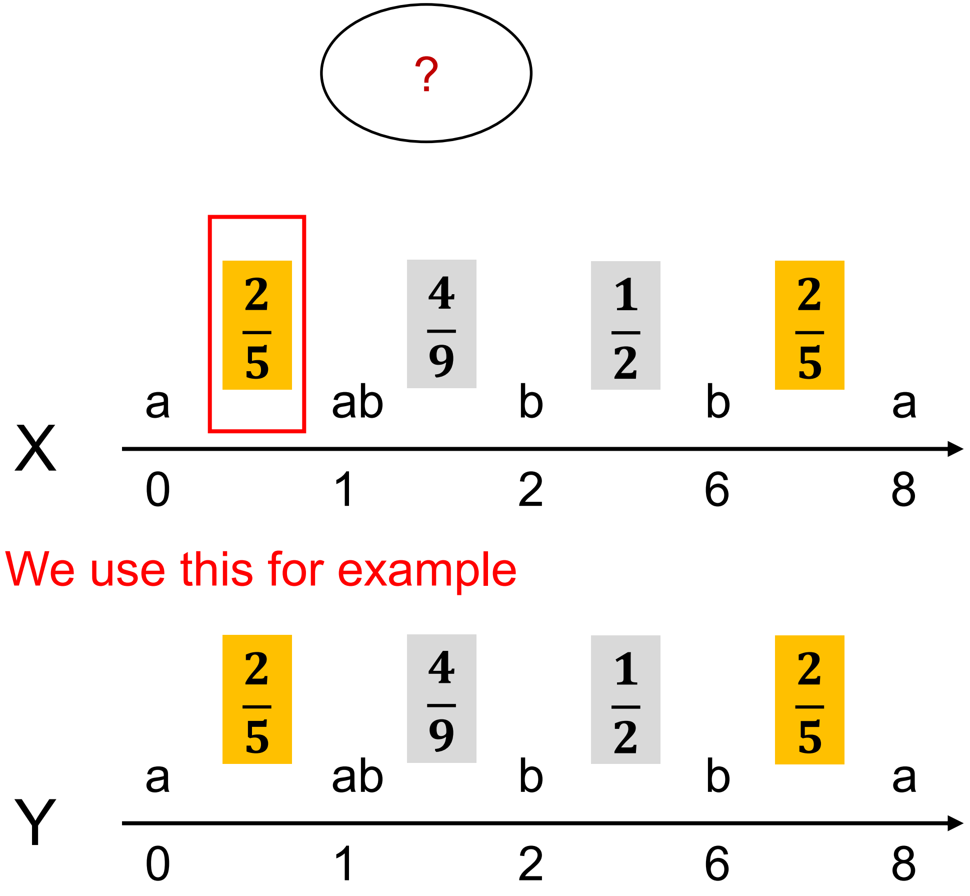
class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6



The same for Y!

QUIZ 01

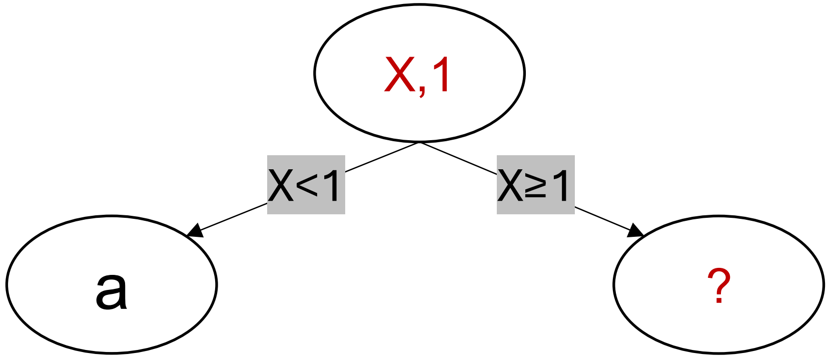
class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6

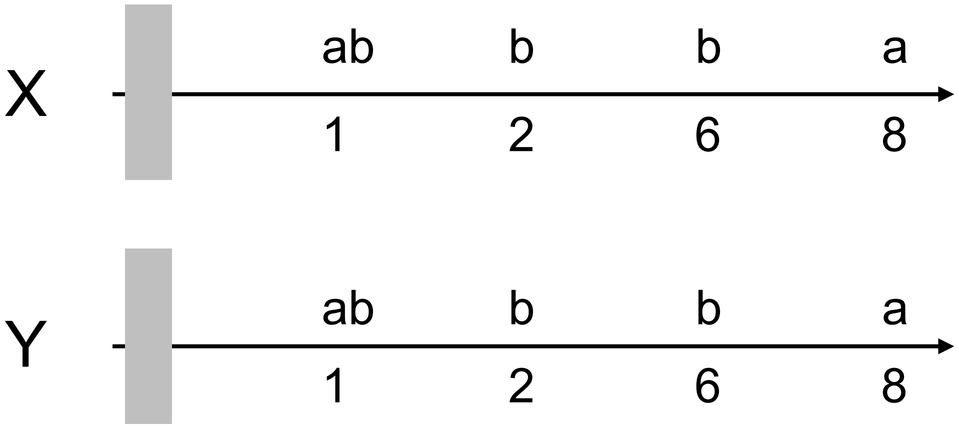


Choose one with the lowest value:

QUIZ 01

class	x	y
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6

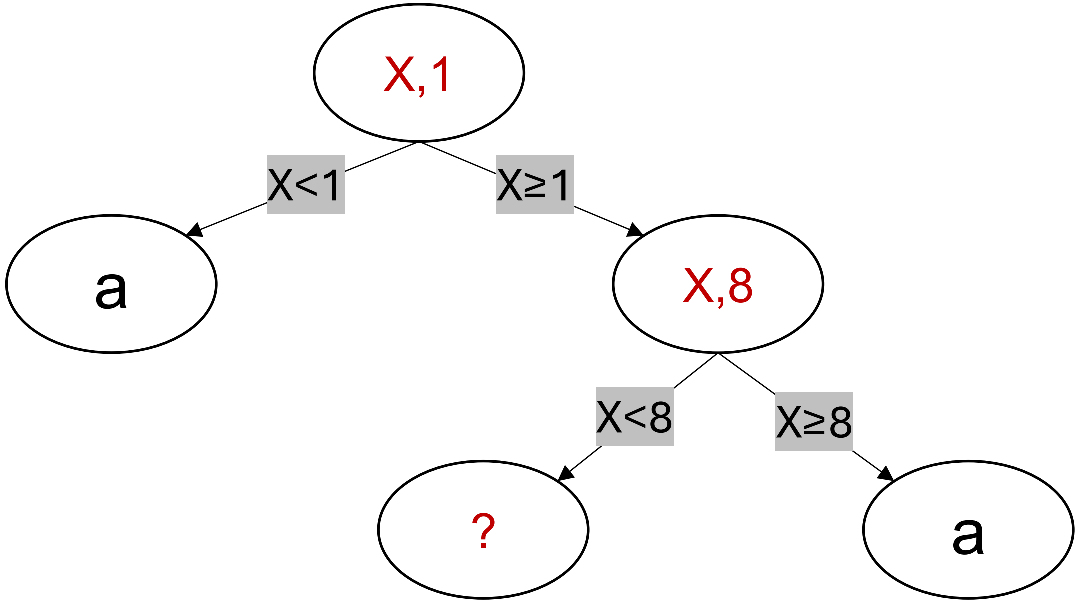




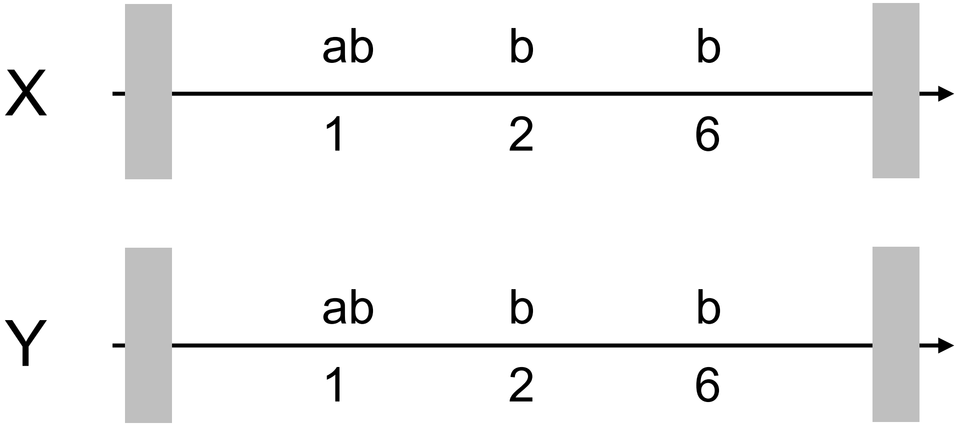
Split the data:

QUIZ 01

class	x	y
a	1	1
b	2	1
b	1	2
b	6	6

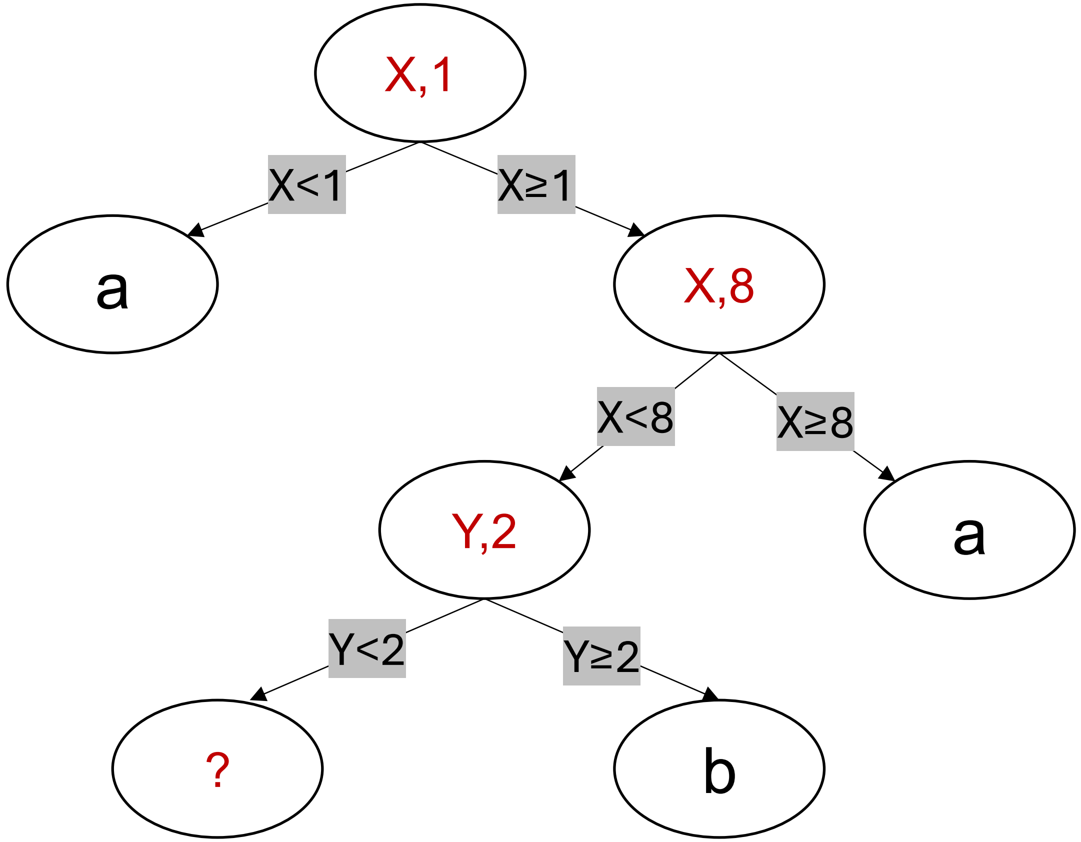


Repeat until done:



QUIZ 01

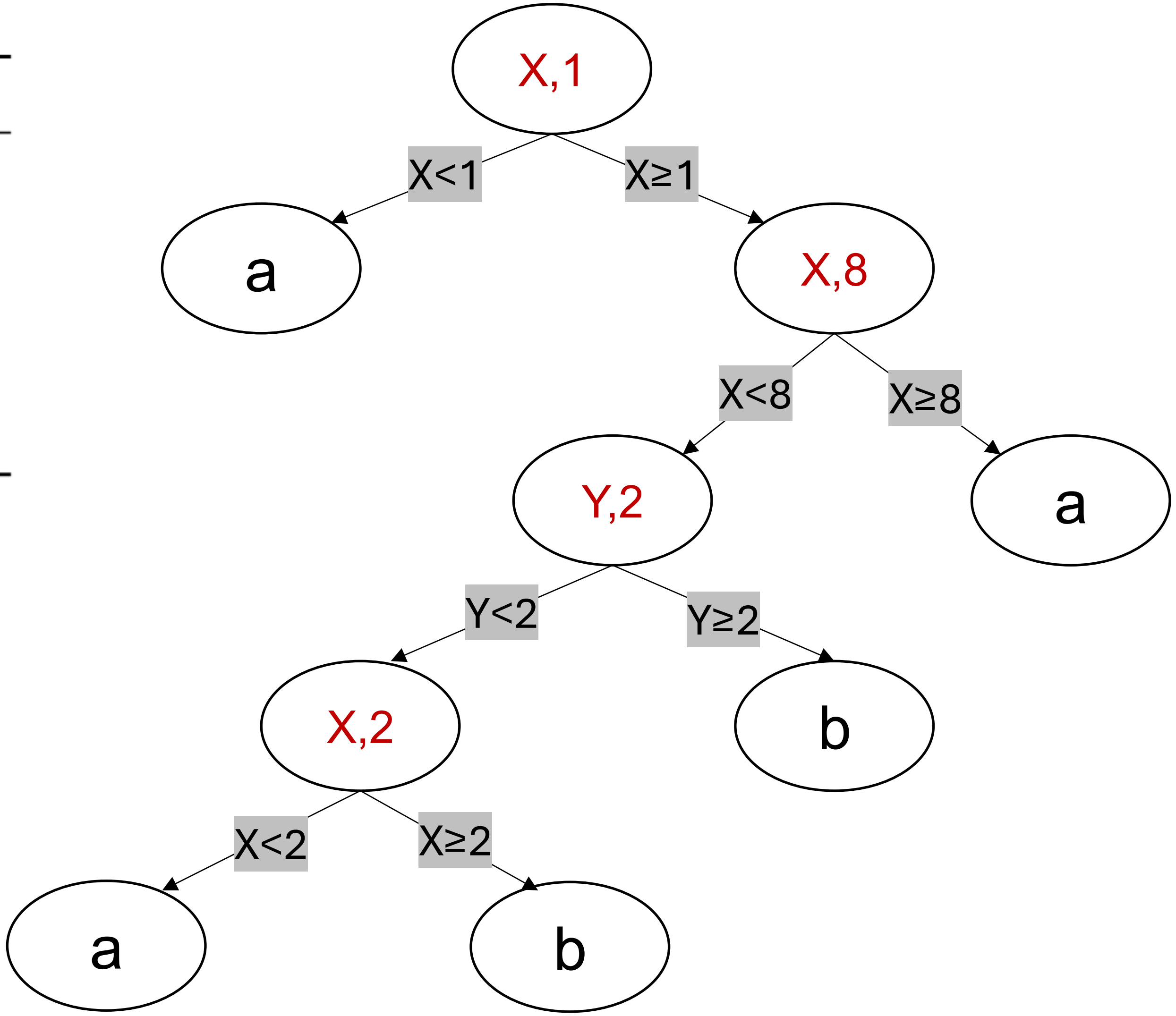
class	x	y
a	1	1
b	2	1



Repeat until done:

QUIZ 01

class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6



An example answer:

(not the only one)

Q&A