Machine Learning – Classification/Clustering

DataBite Summer 2022

Dr. Christan Grant

oudatalab.com

The University of Oklahoma

Data Bite Summer 2022 Schedule

Day 1

Welcome +
Intro to ML

Day 2

Bias and Fairness
+ Introduction to Probability

Day 3

Classifiers and Clustering

Day 4

Model Olympics

From last time…

Introduction to Probability

Rolling a Die Creates a Random Variable

Random Variable →

X	Probability(X)
1	⅙
2	⅙
3	⅙
4	⅙
5	⅙
6	⅙

- If we roll a 6-sided die, what is the probability of rolling a 1?
- What is the probability of rolling an even number?

Die Rolls are Uniform Probabilities

- When we roll a 6-sided die, what is the ”most likely” value?
- Imagine rolling the die 100 times, what would the ”average roll be?

- 3.5 = 1*(1/6) + 2*(1/6) + 3*(1/6) + ... + 6*(1/6)
- 3.5 = (1/100)*(100*1*(1/6) + 100*2*(1/6) + ... + 100*6*(1/6))

- The expected value of a random variable can be thought of as the *mean* or *average*.

```python
import random

rolls = [random.randint(1,6) for i in
range(0,100000)]

average_rolls = sum(rolls)/len(rolls)

print(average_rolls)
```

most

ould the

(1/6)

) + … +

can be

Relationships Among Random Variables

- **Independent variables:** knowing one event has happened does not change the probability that the other happens
  - Probability of rolling a 1 and flipping a head
  - When X and Y are independent, **P(X and Y) = P(X)P(Y)**

- **Dependent variables**: knowing one event has happened gives us new information, affecting the probability that the other happens
  - Probability that the sum of two die rolls being a 5, if the first roll was a 3

Conditional Probability

The probability of X given Y has occurred is P(X|Y), for example,

P(sum = 5 first die = 3) = P(sum is 5 if first die is 3) = 1/6

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Joint Probability

Conditional Probability

Marginal Probability

Probability: Example

- By looking at a table of all possibilities, we found that

$$P(\text{sum} = 5 | \text{first die} = 3) = \frac{1}{6}$$

- Now, using $P(X|Y) = \frac{P(X \text{ and } Y)}{P(Y)}$, we can calculate this without needing to find every possible value of two dice rolls:

$$P(\text{sum} = 5 | \text{ first die} = 3) = \frac{P(\text{second die} = 2 \text{ and first die} = 3)}{P(\text{first die} = 3)}$$

$$= \frac{1/36}{1/6}$$

$$= \frac{1}{6}$$

Conditional Probabilities are not Joint Probabilities

If we let X be the first die roll of value 3, and Y be the second of value 2, and Z be the sum, then

- Conditional probability:

  $P(Z|X) \longrightarrow P(\text{sum} = 5|\text{first die} = 3) = \frac{1}{6}$

- Joint probability:

  $P(X \text{ and } Y) \longrightarrow P(\text{roll 2 and 3}) = \frac{2}{36}$

- Probability of Z

  $P(Z) = P(X + Y) \longrightarrow P(\text{sum} = 5) = \frac{4}{36}$

Conditional, Joint, & Marginal Probabilities are Related

Let X and Y be random variables.

1. If X and Y are independent then $P(X|Y) = \frac{P(X \text{ and } Y)}{P(Y)} = \frac{P(X)P(Y)}{P(Y)} = P(X)$

   Example: let X be the outcome of rolling a 6-sided die and let Y be the outcome of flipping a coin. Suppose we know that Y is "heads." What is the probability that we roll a 3? This tells us that $P(\textbf{roll 3}|\textbf{heads}) = P(\textbf{roll 3})$. *This matches intuition — flipping a coin does not change the outcome of rolling a die.*

Conditional, Joint, & Marginal Probabilities are Related

Let X and Y be random variables.

2. $P(X) = \sum_{Y} P(X|Y)P(Y)$

Example: let X be the sum of rolling two dice and let Y be the outcome of the first die roll. If we want to know the probability that X is 3, then this tells us $P(\text{sum is 3}) = \sum_{Y} P(\text{sum is 3}|\text{first roll was Y})P(\text{first roll was Y})$*. From here, we know that the sum can never be 3 unless the first roll is 1 or 2. Thus,*

$$P(\text{sum is 3}) = P(\text{sum is 3}|\text{first roll was 1})P(\text{first roll was 1}) + P(\text{sum is 3}|\text{first roll was 2})P(\text{first roll was 2})$$
$$= \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{2}{36}$$

Conditional, Joint, & Marginal Probabilities are Related

Let X and Y be random variables.

3. $P(X) = \sum_{Y} P(X \text{ and } Y)$

Example: let X be the outcome of a first die roll and let Y be the outcome of a second die roll. If we want to find the probability that X is 3, this tells us that

$$P(X=3) = \sum_{y=1}^{6} P(X=3 \text{ and } Y=y)$$
$$= \sum_{y=1}^{6} \frac{1}{36} = \frac{6}{36} = \frac{1}{6}$$

Conditional, Joint, & Marginal Probabilities are Related

Let X and Y be random variables.

1. If X and Y are independent then $P(X|Y) = \frac{P(X \text{ and } Y)}{P(Y)} = \frac{P(X)P(Y)}{P(Y)} = P(X)$
2. $P(X) = \sum_{Y} P(X|Y)P(Y)$
3. $P(X) = \sum_{Y} P(X \text{ and } Y)$

Conditional Probability

Joint Probability

Marginal Probability

Conditional Probabilities and **Bayes' Theorem**

Sometimes we want to find P(X|Y) when we already know P(Y | X)

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Bayes' Theorem: Example

Sometimes we want to find $P(X|Y)$ when we already know $P(Y|X)$.

For instance, $P(\text{first die} = 3|\text{sum} = 5) = \frac{1}{4}$.

We can verify this using Bayes' Theorem.

$$\begin{aligned} P(\text{first die} = 3|\text{sum} = 5) &= \frac{P(\text{sum} = 5|\text{first die} = 3)P(\text{first die} = 3)}{P(\text{sum} = 5)} \\ &= \frac{\frac{1}{6} \cdot \frac{1}{6}}{\frac{4}{36}} \\ &= \frac{1}{4} \end{aligned}$$

Sample Exercise: Peanut Chocolate Detector

Assumptions: Suppose we have a new device that distinguishes whether or not a type of chocolate contains peanuts. If a chocolate contains peanuts, 99% of the time it correctly reports a positive result. Likewise, if a chocolate does not contain peanuts, 99% of the time it correctly reports a negative result. Imagine 1% of all chocolates contain peanuts.

Question: If the device reports that a chocolate contains peanuts, what is the probability that the chocolate *actually does* contain peanuts?

Sample Exercise: Peanut Chocolate Detector

p = random variable indicating whether peanuts are in a chocolate bar

d = random variable indicating whether we detected peanuts in a chocolate bar.

We were given $P(p) = 0.01$, $P(d|p) = 0.99$, and $P(\text{not } d|\text{not } p) = 0.99$.

We can then calculate $P(d|\text{not } p) = 0.01$ and$P(\text{not } p) = 0.01$.

Then, $P(d) = P(d|p)P(p) + P(\text{not } d|\text{not } p)P(\text{not } p)$.

By Bayes' Theorem, $P(p|d) = \frac{P(d|p)P(p)}{P(d)} = \frac{0.99 \cdot 0.01}{0.0198} = 0.5$.

Clustering and Classification

Definitions

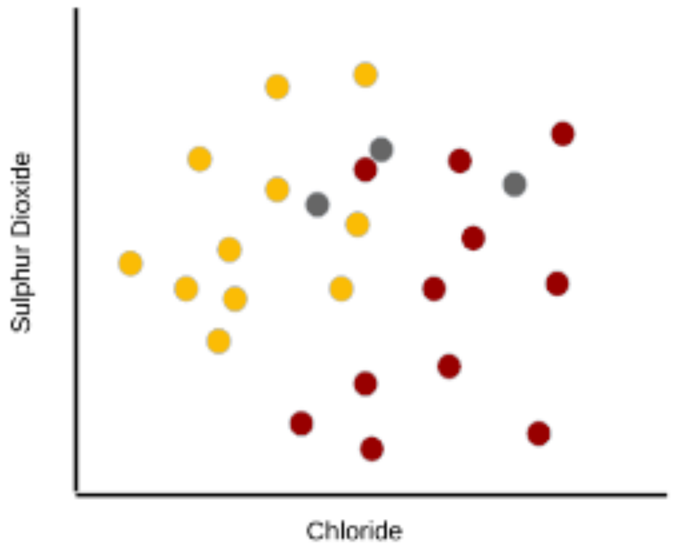
- Classification
  - “the action or process of classifying something according to shared qualities or characteristics.” (https://languages.oup.com/)
- Clustering
  - "Cluster analysis or clustering is the task of grouping a set of objects so that objects in the same group are more similar to each other than to those in other groups." en.wikipedia.org/wiki/Cluster_analysis
- Regression
  - Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship (investopedia.com)

```
[
  'tiger': 0.96456,
  'lion': 0.7456
  'cougar': 0.6789,
  'mountain_lion': 0.5467,
  'lynx': 0.5326,
  ...
]
```

Classifying Red and White Drinks

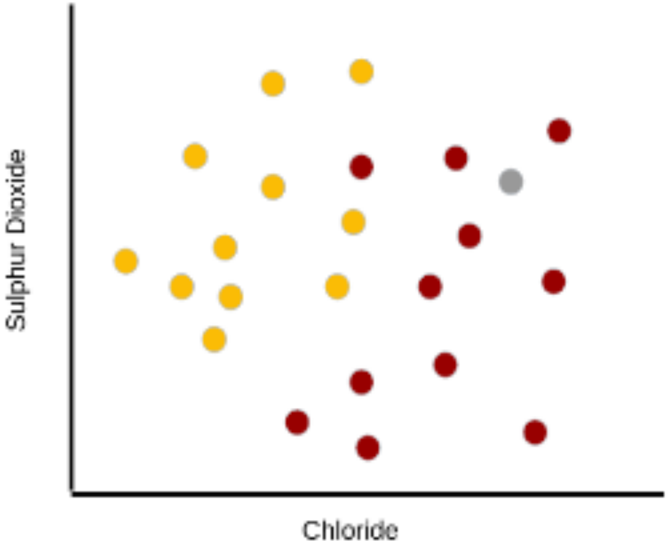
Classifying Red and White Drinks

- We will use a technique called **K Nearest Neighbors**
- This technique says “what other examples” are similar.
- The term K refers to the number of examples that are you compare against.



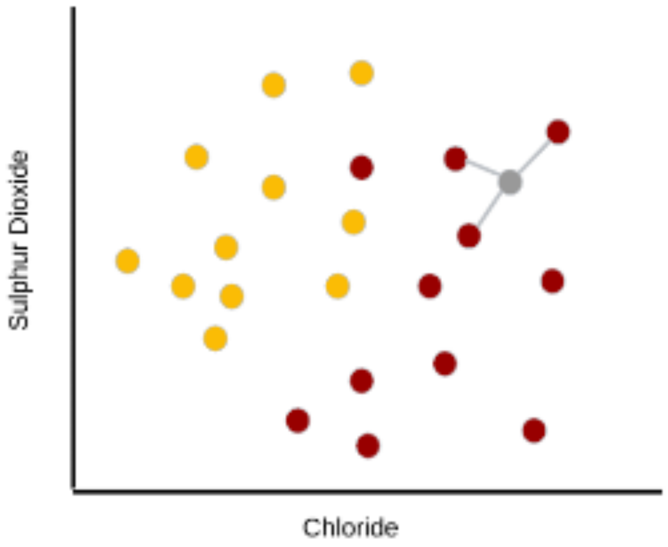
How do we classify a new point?

- K = 3
- K=5



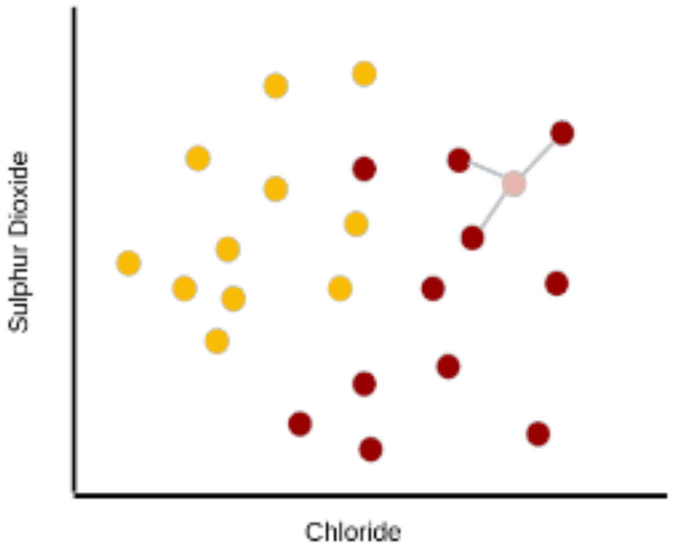
How do we classify a new point?

- K = 3



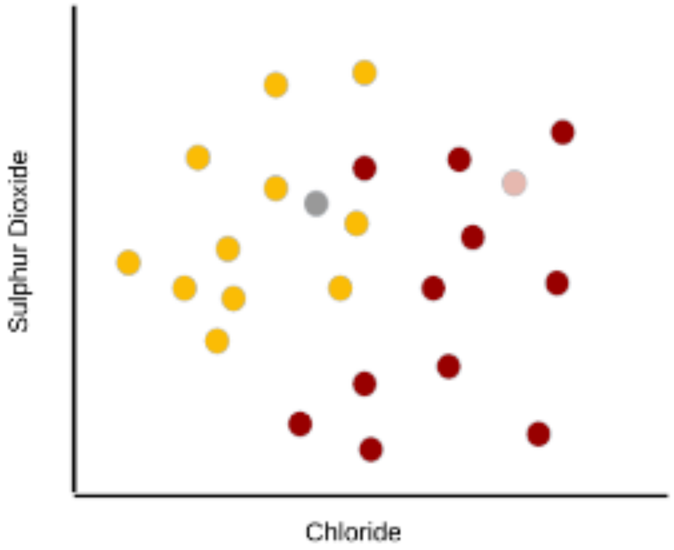
How do we classify a new point?

- K = 3



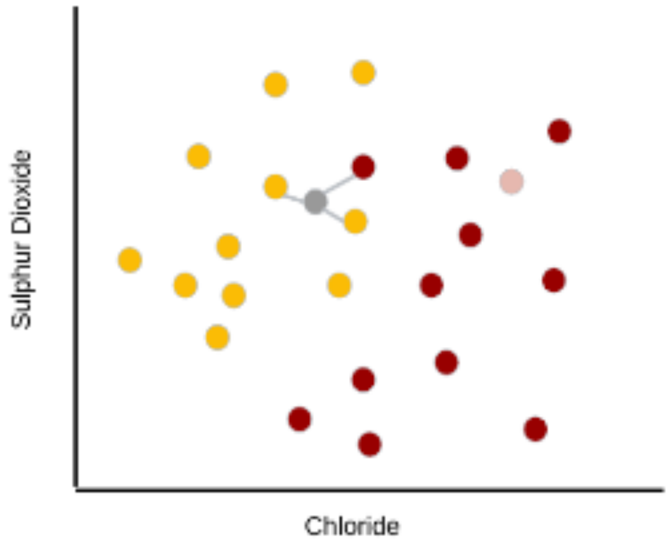
How do we classify a new point?

- K = 3



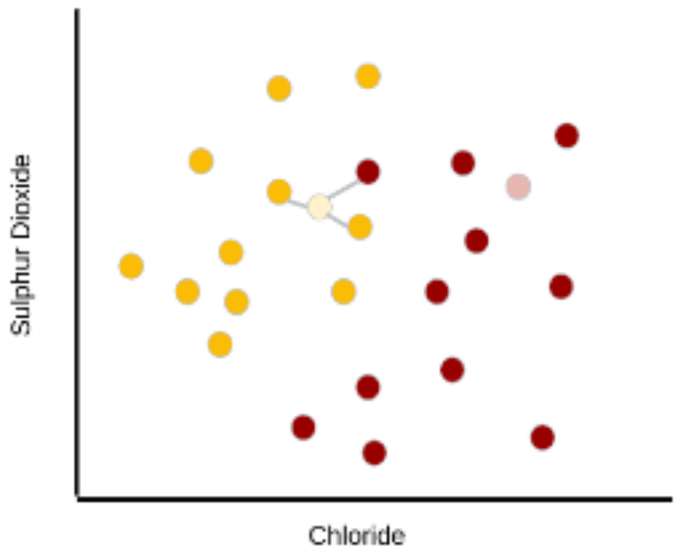
How do we classify a new point?

- K = 3



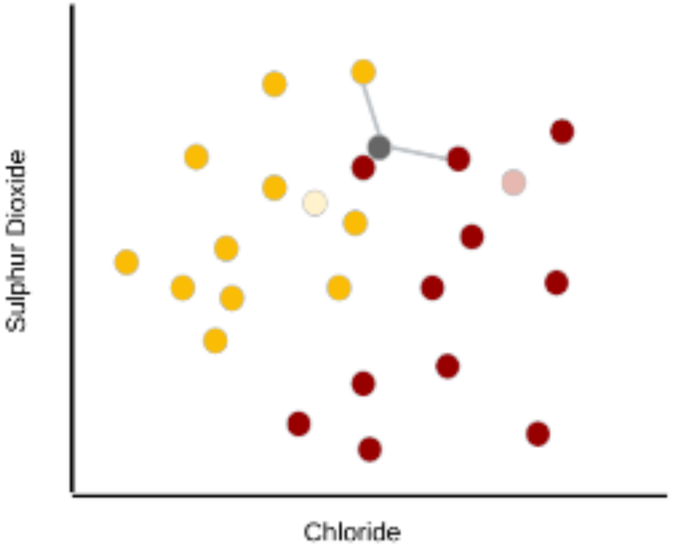
How do we classify a new point?

- K = 3



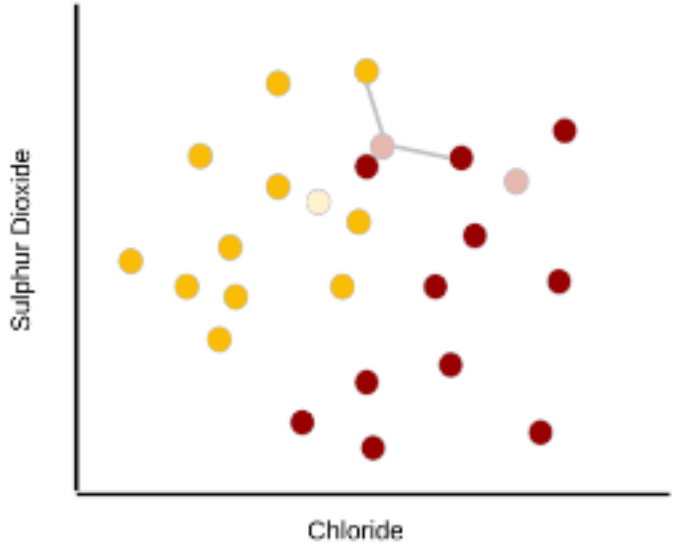
How do we classify a new point?

- K = 3



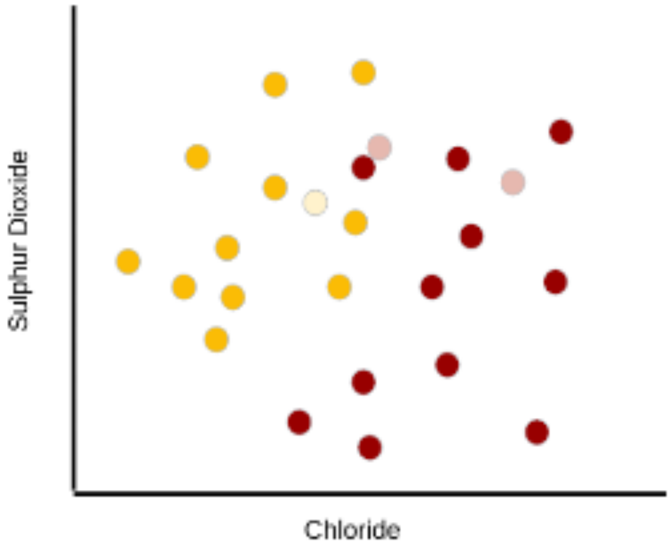
How do we classify a new point?

- K = 3



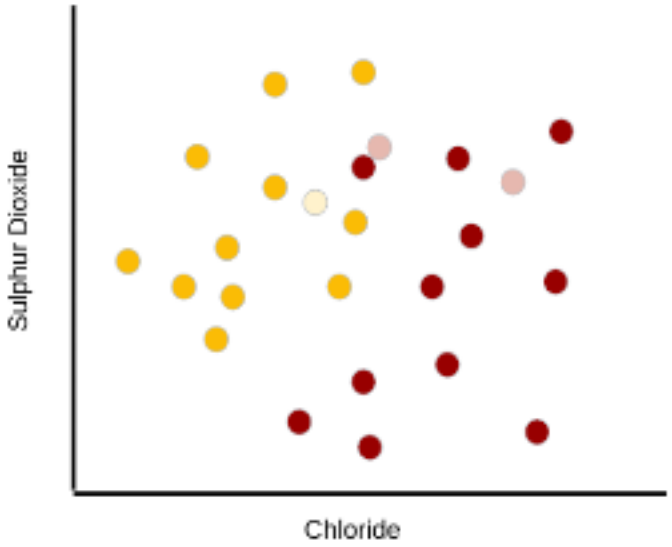
How do we classify a new point?

- K = 3



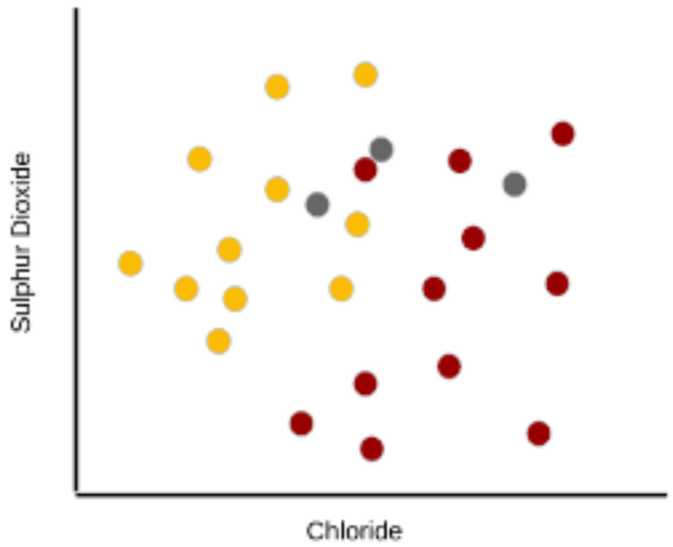
How do we classify a new point?

- K = 3



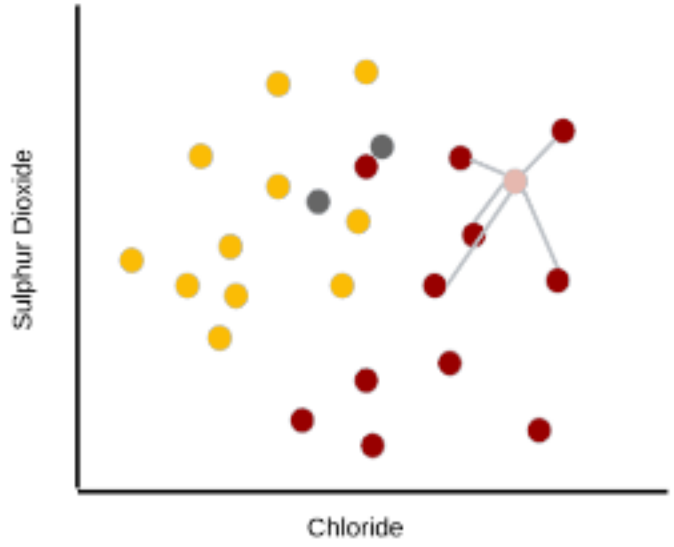
How do we classify a new point?

- K = 3



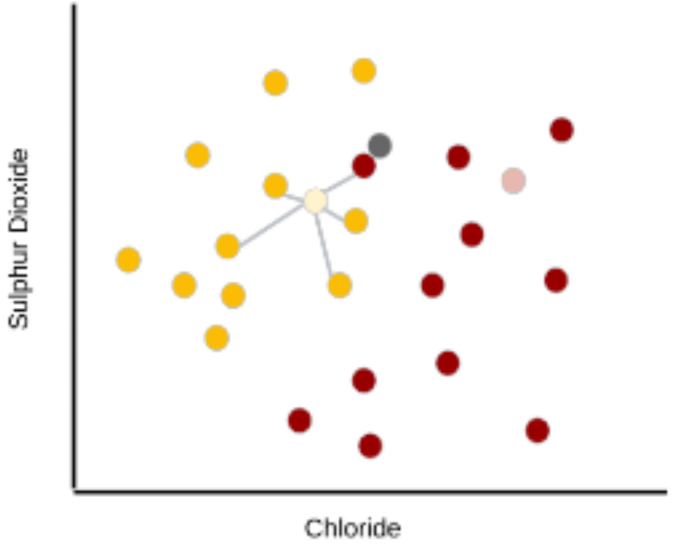
How do we classify a new point?

- K = 3
- K=5



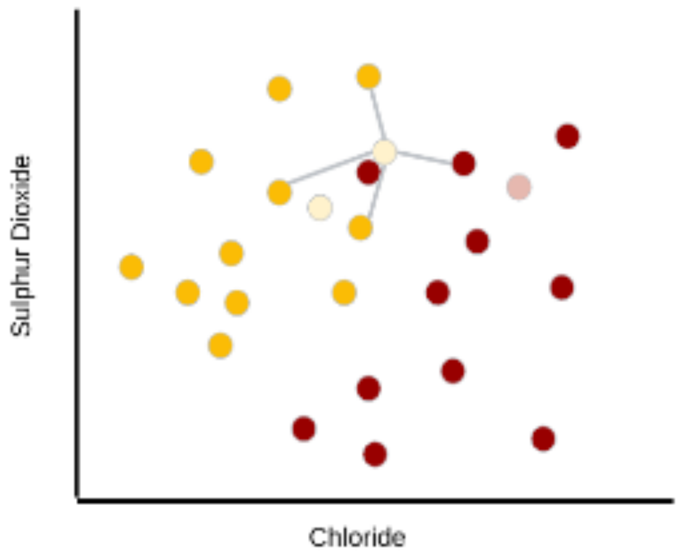
How do we classify a new point?

- K = 3
- K=5



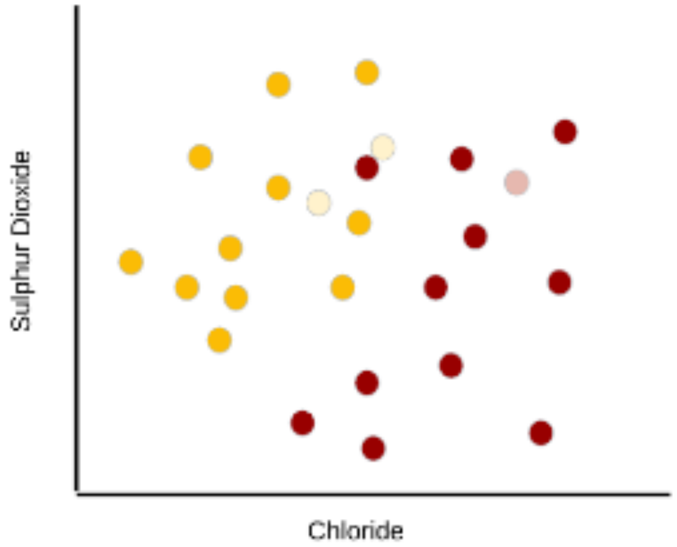
How do we classify a new point?

- K = 3
- K=5



How do we classify a new point?

- K = 3
- K=5

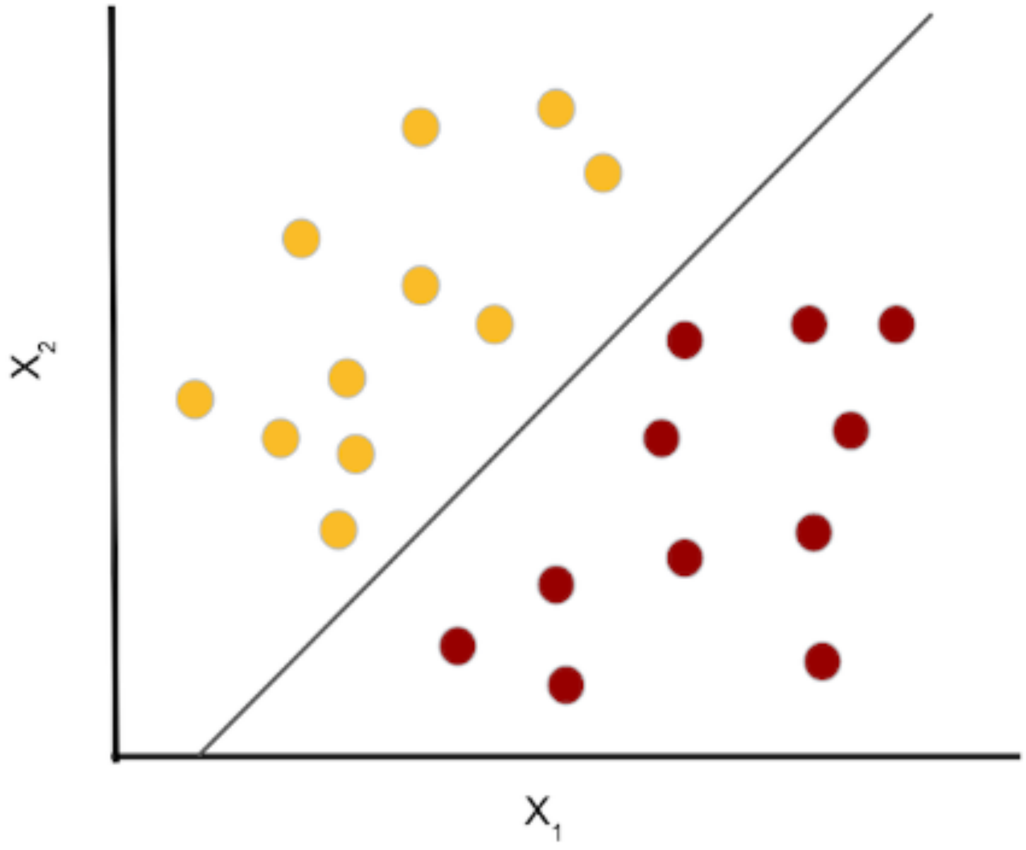


How do we classify a new point?

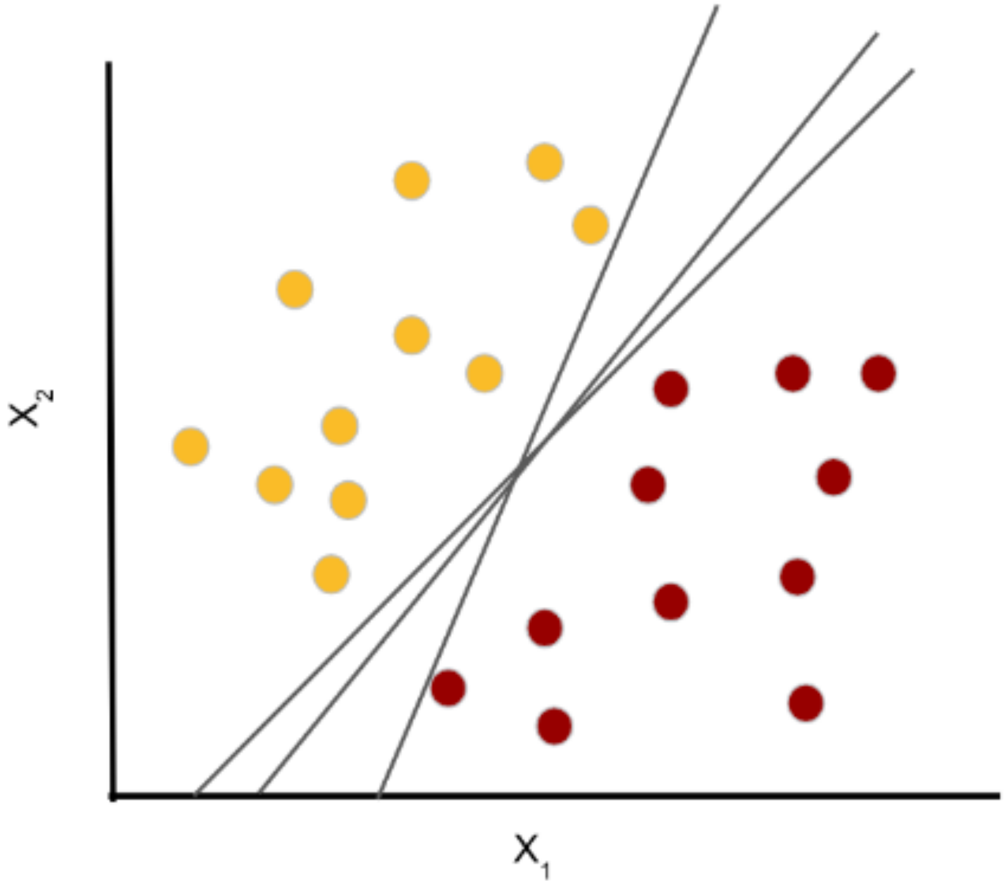
- K = 3
- K=5

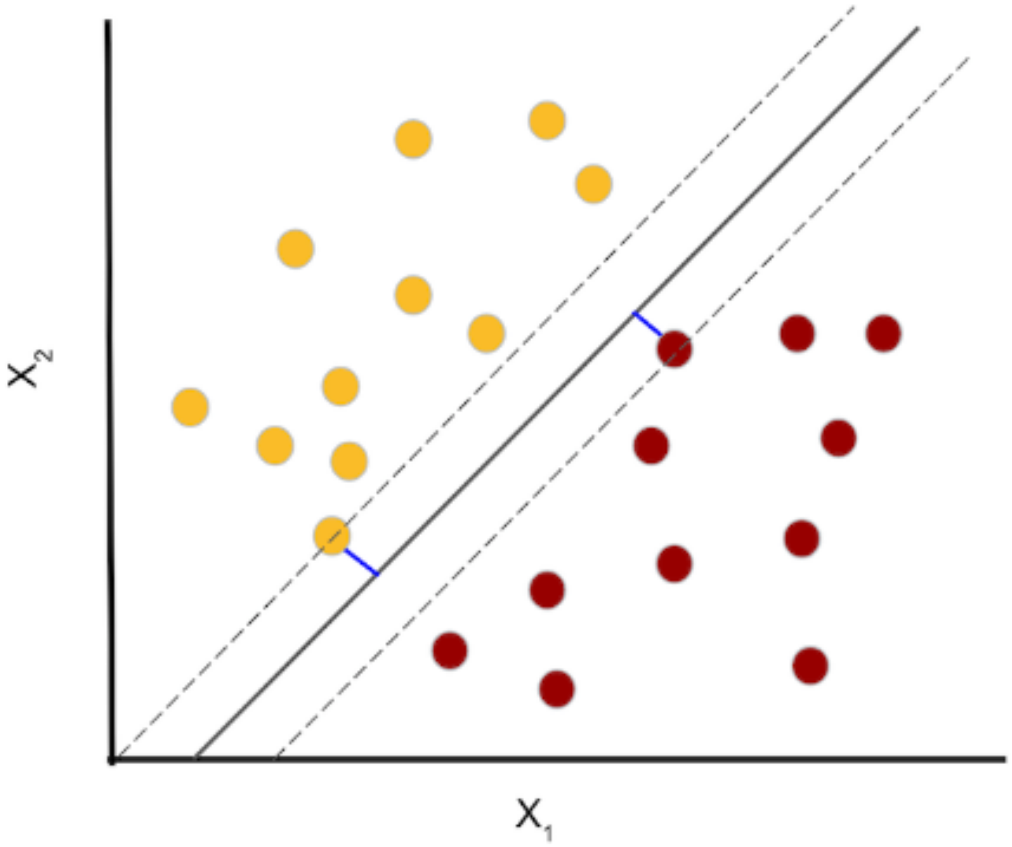
Classifying Red and White Drinks

- Another technique is called Support Vector Machines (SVM)
- This techniques tries to identify a dividing line between instances. (Or the vectors/lines that separate the points)
- New points are classified by where they call on the line

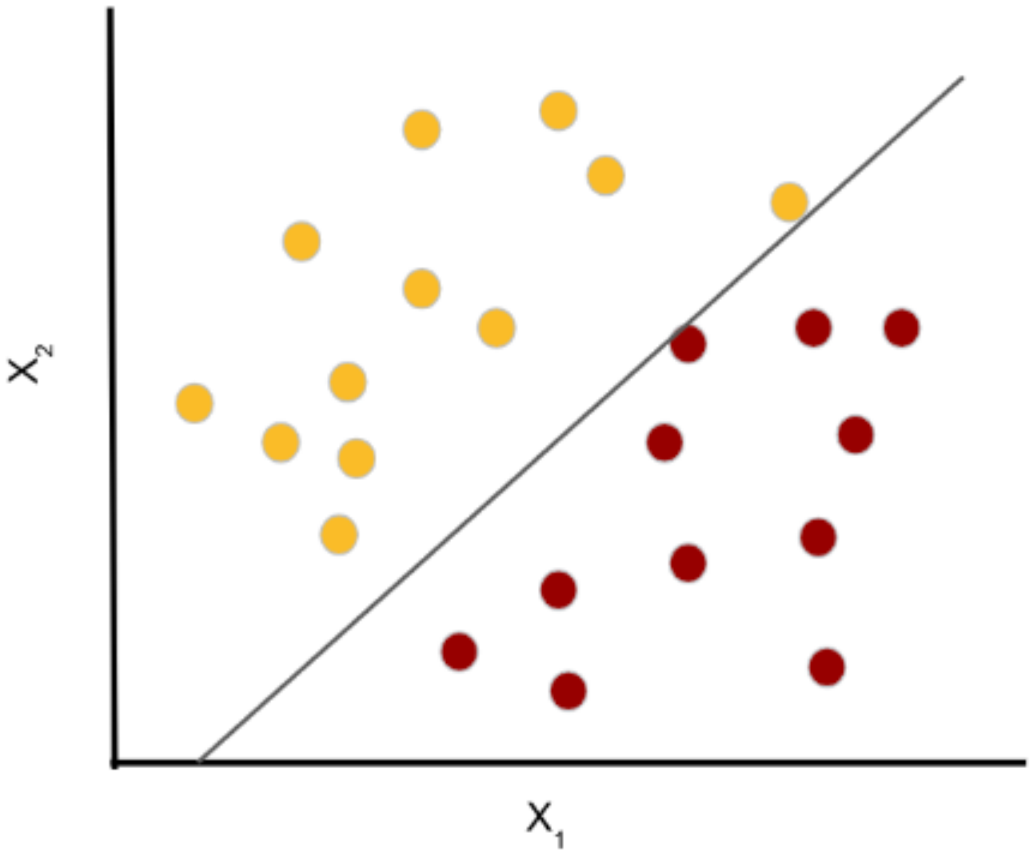


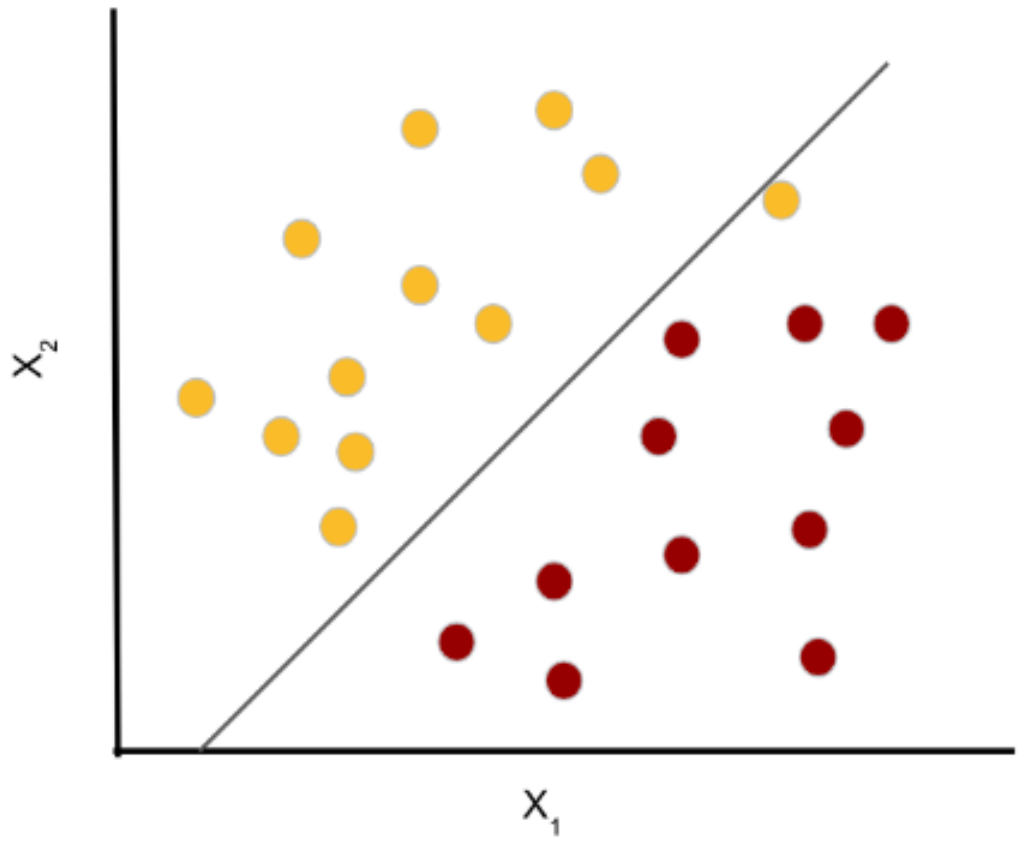
X_2
X_1

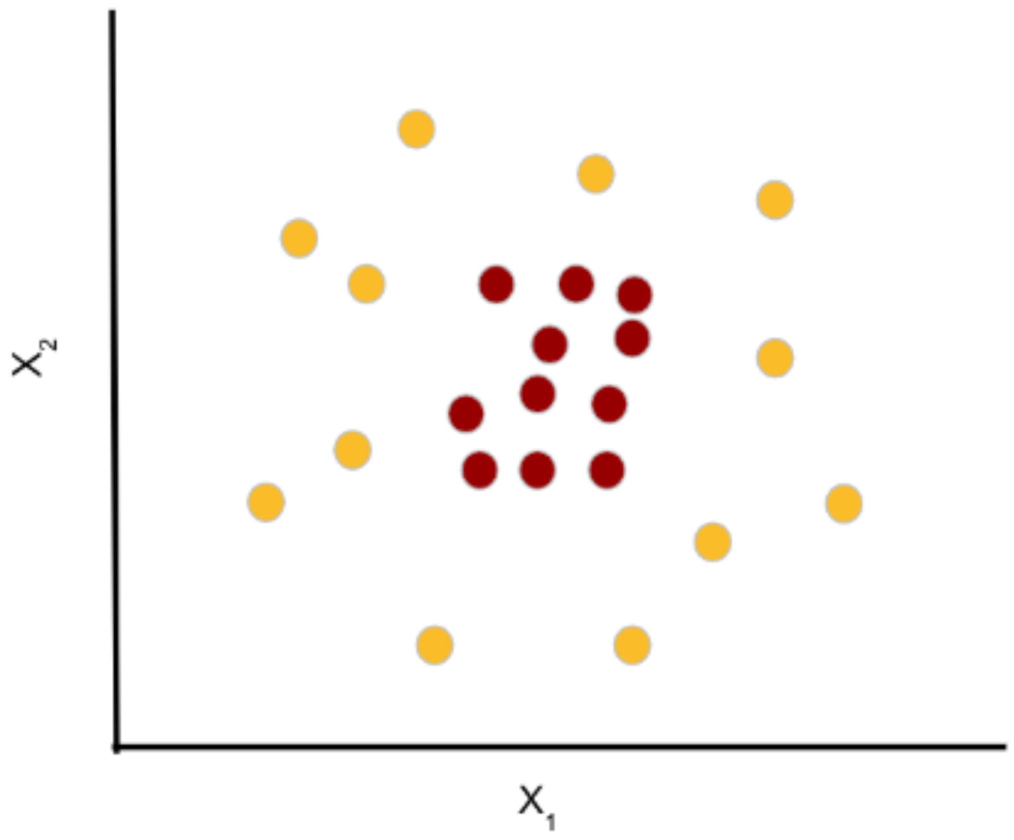

X
2
X
1

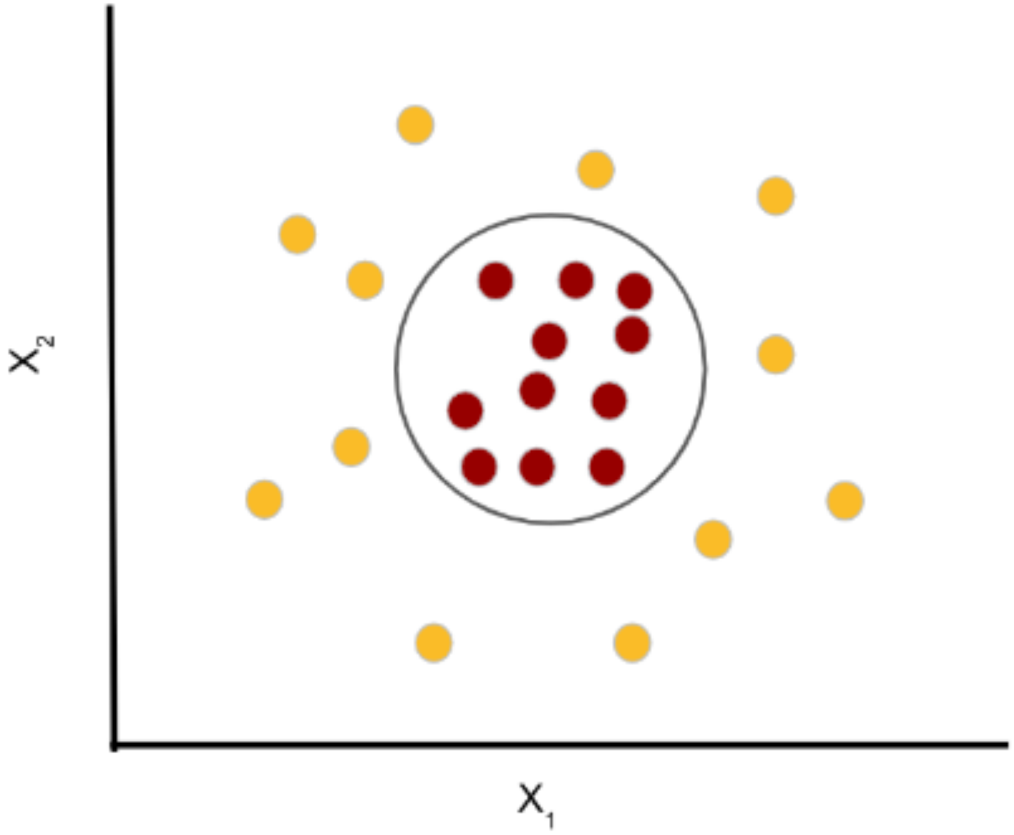


X_2
X_1


X2
X1


X2
X1

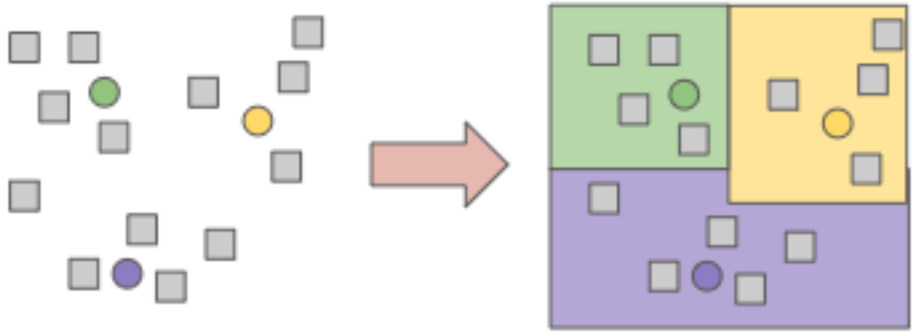

X_2
X_1

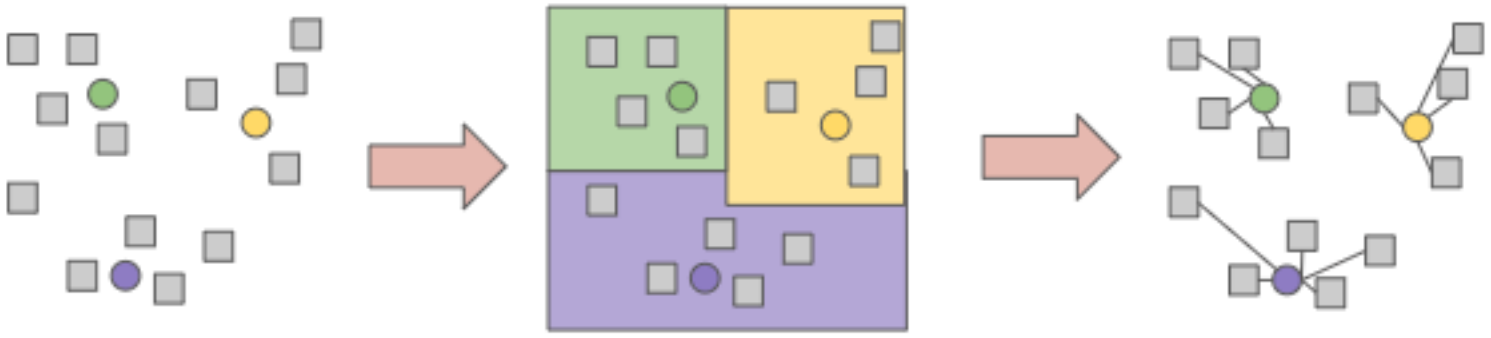

X2
X1

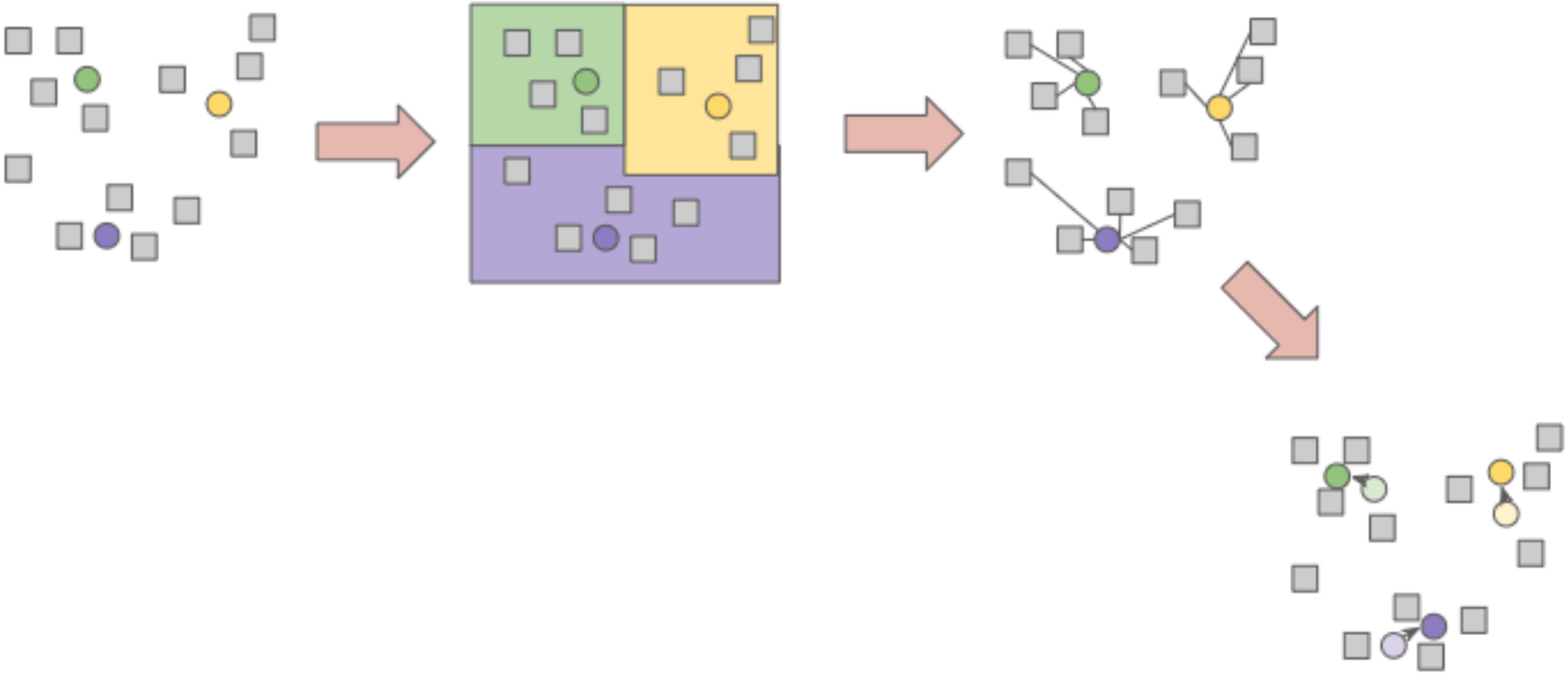
Clustering

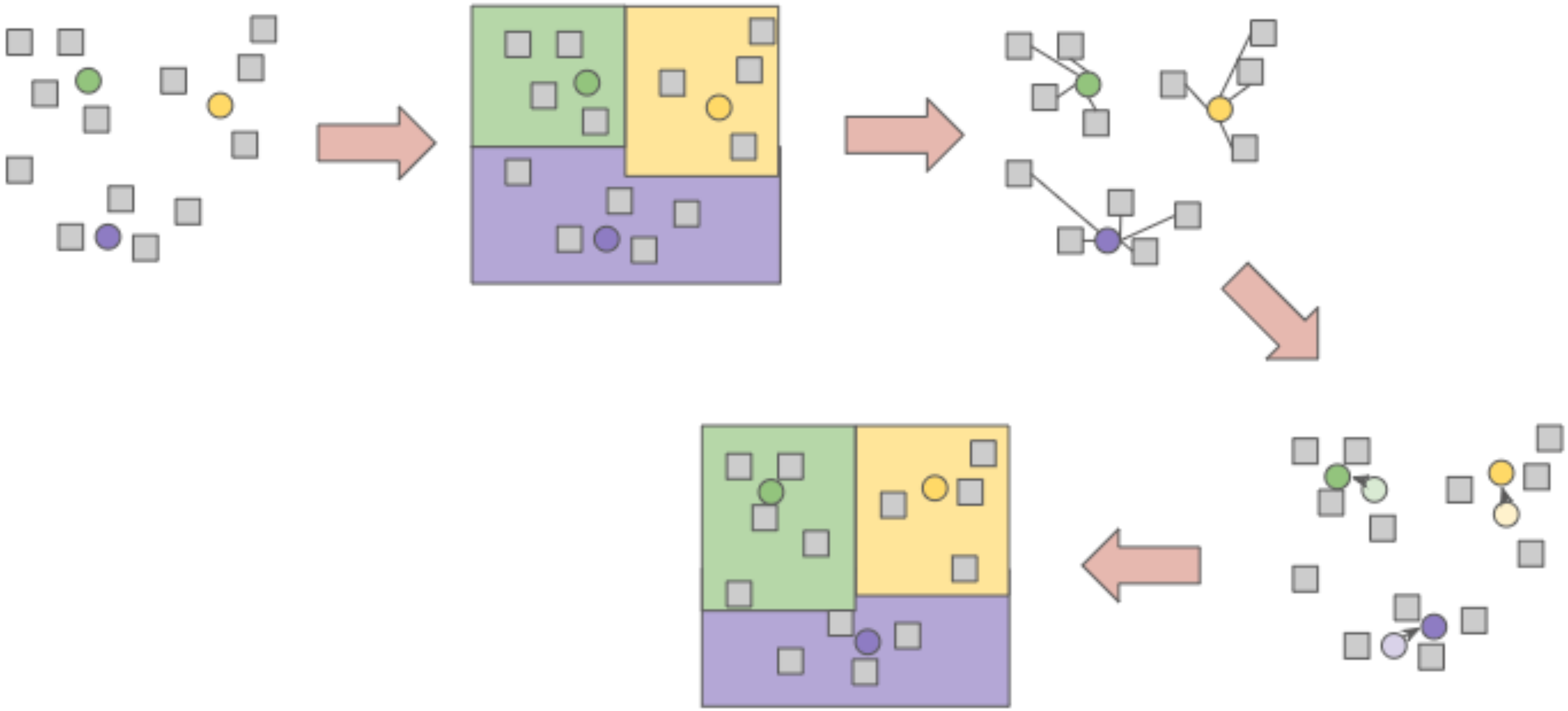
K-Means

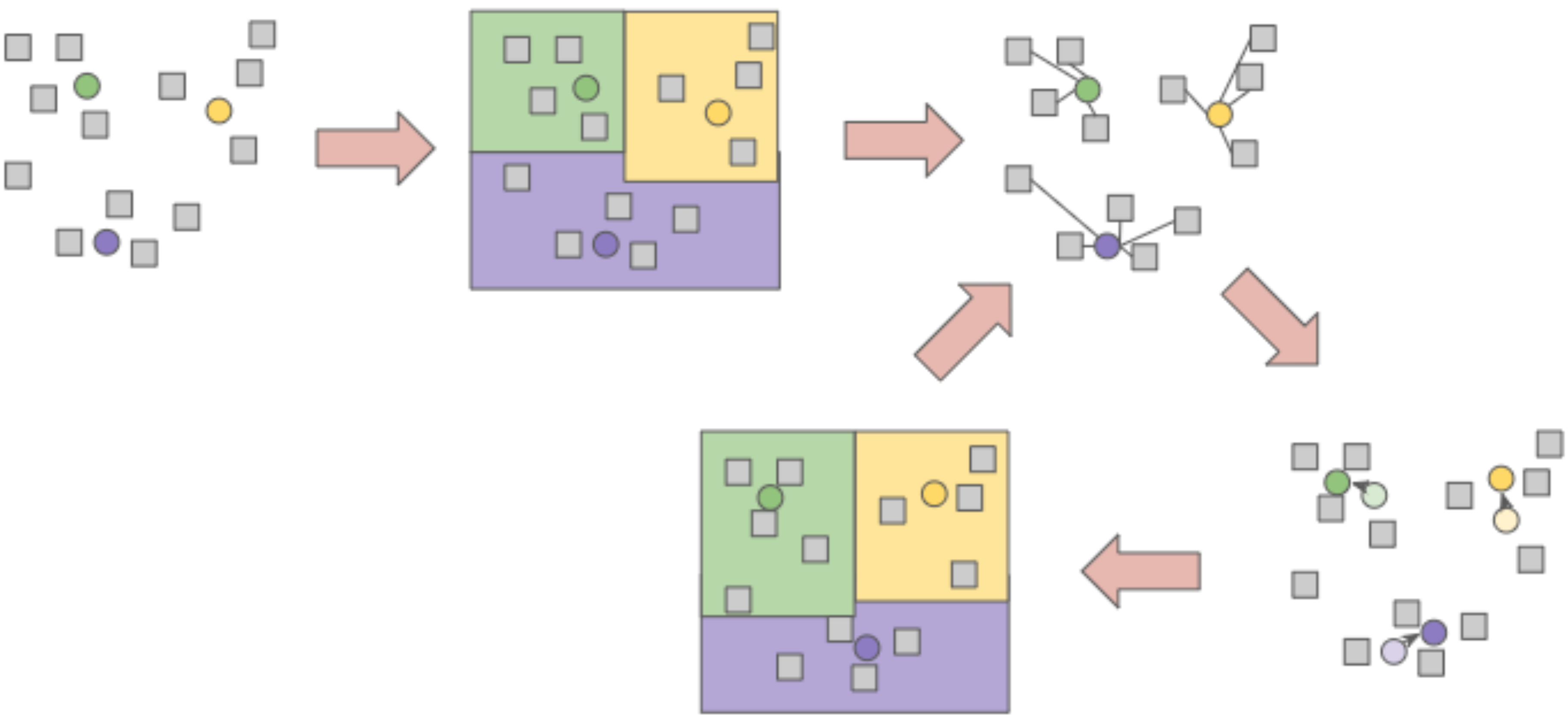
- K-Means uses K important points to identify clusters.
- These centroids area updated and the points are relabeled and reassigned clusters.
- This process repeats until the clusters no longer change.











Thanks!

@oudatalab