

Introduction to Artificial Intelligence

English practicals 10: Machine learning

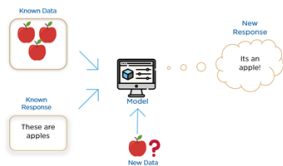
Marika Ivanová

Department of Theoretical Computer Science and Mathematical Logic (KTIML)
Faculty of Mathematics and Physics

May 13th 2021

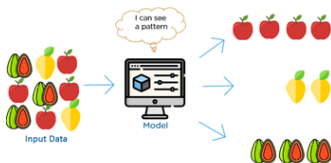
Machine learning

Supervised (task driven)



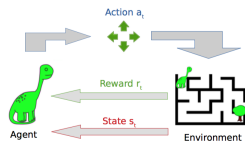
source: towardsdatascience.com

Unsupervised (data driven)



source: towardsdatascience.com

Reinforcement (learn from mistakes)

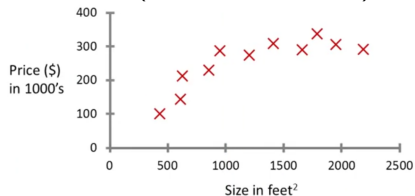


source: towardsdatascience.com

- Ad popularity
- Spam classification
- Face recognition
- We already touched upon reinforcement learning when last time we talked about the robot moving in the maze
- Recommendation systems
- Buying habits
- Grouping user logs
- Video games
- Industrial simulations
- Robot navigation

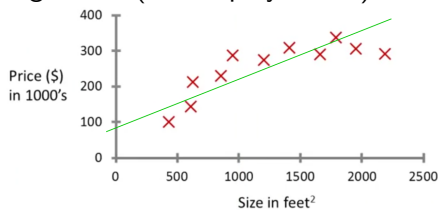
Supervised learning: regression vs classification

Regression (linear, polynomial)



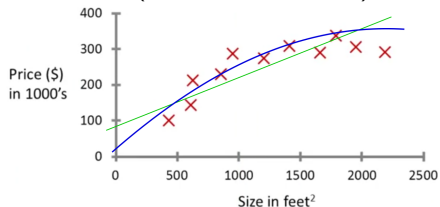
Supervised learning: regression vs classification

Regression (linear, polynomial)



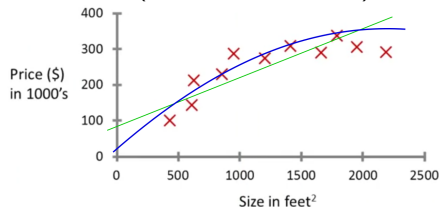
Supervised learning: regression vs classification

Regression (linear, polynomial)



Supervised learning: regression vs classification

Regression (linear, polynomial)

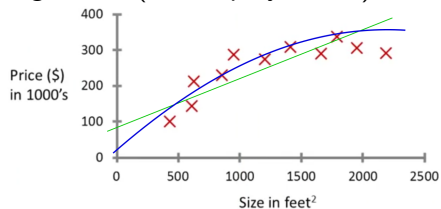


Classification

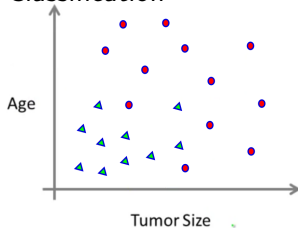


Supervised learning: regression vs classification

Regression (linear, polynomial)

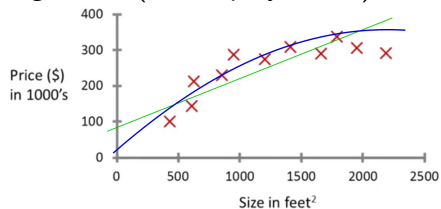


Classification

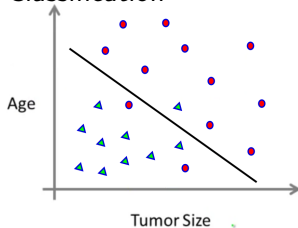


Supervised learning: regression vs classification

Regression (linear, polynomial)



Classification



Exercise: regression vs classification

Decide what type of problems are the following tasks:

- ① You have an inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
 - ② You'd like to examine individual customer accounts, and for each account decide if it has been hacked.
-
- a) Treat both as classification problems
 - b) Treat problem (1) as classification, problem (2) as regression
 - c) Treat problem (1) as regression, problem (2) as classification
 - d) Treat both as regression problems

Exercise: regression vs classification

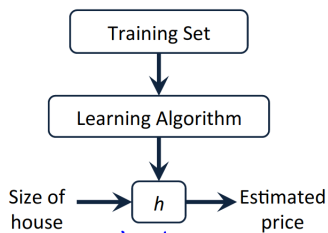
Decide what type of problems are the following tasks:

- ① You have an inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
- ② You'd like to examine individual customer accounts, and for each account decide if it has been hacked.

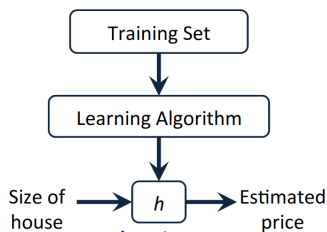
- a) Treat both as classification problems
- b) Treat problem (1) as classification, problem (2) as regression
- c) Treat problem (1) as regression, problem (2) as classification
- d) Treat both as regression problems

C is correct

Linear regression

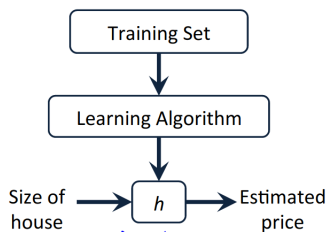


Linear regression

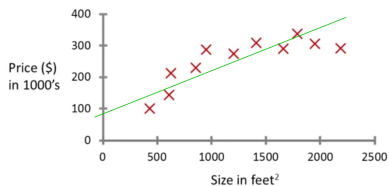


Size in feet ² (x)	Price (\$) in 1000's (y)
2104	460
1416	232
1534	315
852	178
...	...

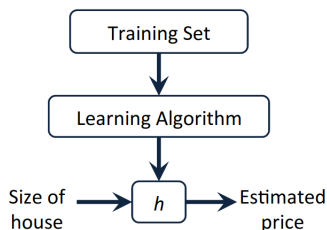
Linear regression



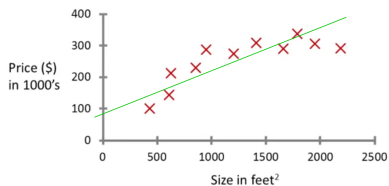
Size in feet ² (x)	Price (\$) in 1000's (y)
2104	460
1416	232
1534	315
852	178
...	...



Linear regression



Size in feet ² (x)	Price (\$) in 1000's (y)
2104	460
1416	232
1534	315
852	178
...	...

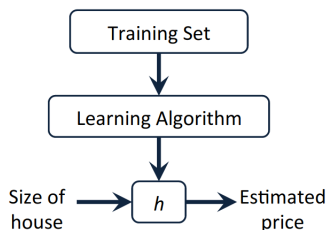


Hypothesis:

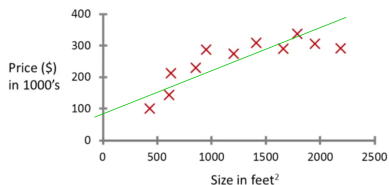
$$h_w(x) = w_0 + w_1(x)$$

$w_0, w_1 \dots$ parameters
how to choose w ?

Linear regression



Size in feet ² (x)	Price (\$) in 1000's (y)
2104	460
1416	232
1534	315
852	178
...	...



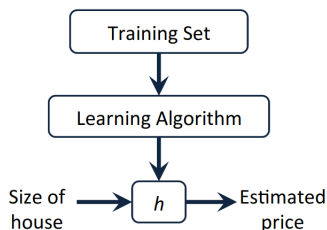
Hypothesis:

$$h_w(x) = w_0 + w_1(x)$$

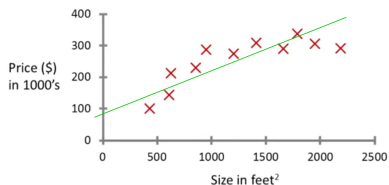
$w_0, w_1 \dots$ parameters
how to choose w ?

Choose w_0, w_1 so that $h_w(x)$ is close to y for training examples (x, y)

Linear regression



Size in feet ² (x)	Price (\$) in 1000's (y)
2104	460
1416	232
1534	315
852	178
...	...



Hypothesis:

$$h_w(x) = w_0 + w_1(x)$$

$w_0, w_1 \dots$ parameters
how to choose w ?

Choose w_0, w_1 so that $h_w(x)$ is close to y for training examples (x, y)

$$\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m \left(h_w(x^{(i)}) - y^{(i)} \right)^2$$

Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$

Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

Linear regression

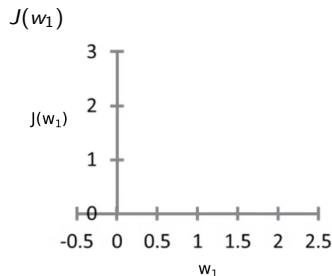
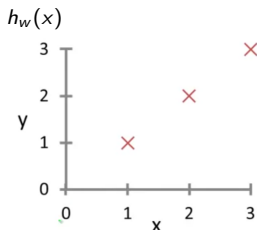
- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

Intuition: simplify the hypothesis: $h_w(x) = w_1 x_1$, we want $\min_{w_1} J(w_1)$

Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

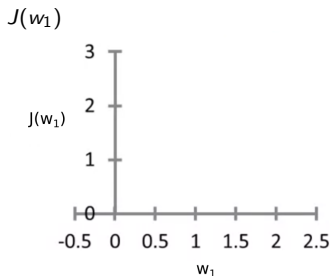
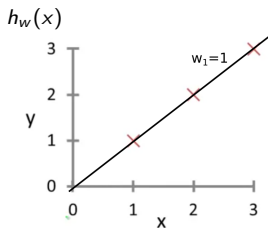
Intuition: simplify the hypothesis: $h_w(x) = w_1 x_1$, we want $\min_{w_1} J(w_1)$



Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

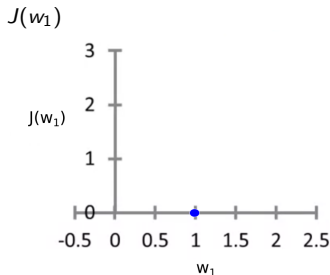
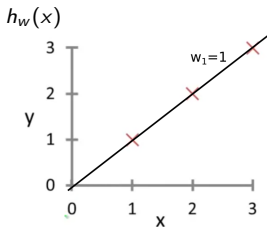
Intuition: simplify the hypothesis: $h_w(x) = w_1 x_1$, we want $\min_{w_1} J(w_1)$



Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

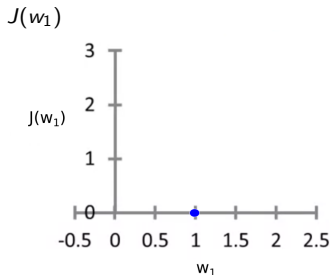
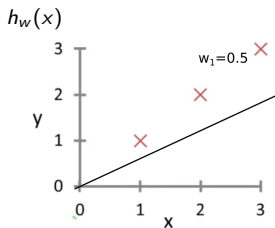
Intuition: simplify the hypothesis: $h_w(x) = w_1 x_1$, we want $\min_{w_1} J(w_1)$



Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

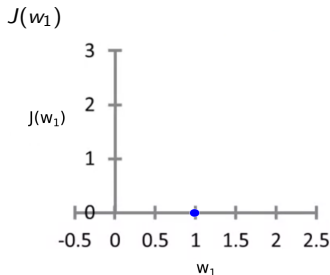
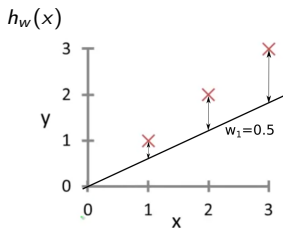
Intuition: simplify the hypothesis: $h_w(x) = w_1 x_1$, we want $\min_{w_1} J(w_1)$



Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

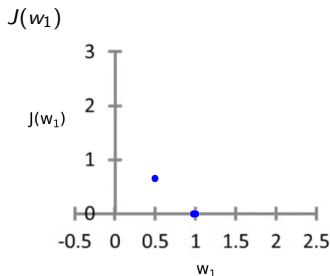
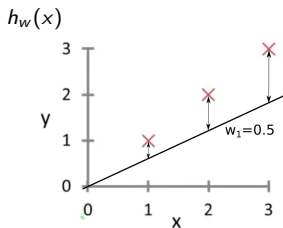
Intuition: simplify the hypothesis: $h_w(x) = w_1 x_1$, we want $\min_{w_1} J(w_1)$



Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

Intuition: simplify the hypothesis: $h_w(x) = w_1 x_1$, we want $\min_{w_1} J(w_1)$

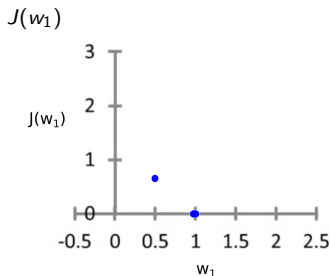
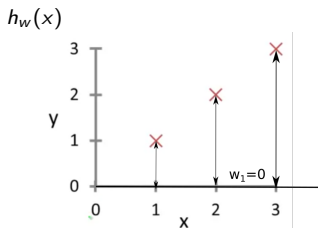


$$J(0.5) = 1/6 * [(0.5 - 1)^2 + (1 - 2)^2 + (1.5 - 3)^2] = 3.5/6 \approx 0.58$$

Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

Intuition: simplify the hypothesis: $h_w(x) = w_1 x_1$, we want $\min_{w_1} J(w_1)$

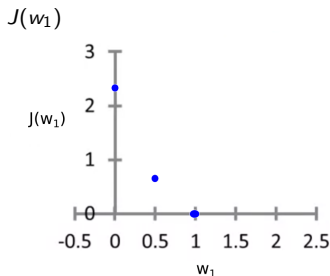
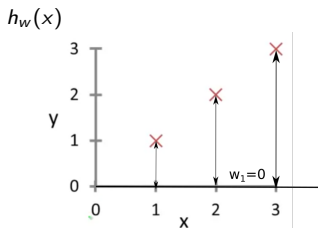


$$J(0.5) = 1/6 * [(0.5 - 1)^2 + (1 - 2)^2 + (1.5 - 3)^2] = 3.5/6 \approx 0.58$$

Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

Intuition: simplify the hypothesis: $h_w(x) = w_1 x_1$, we want $\min_{w_1} J(w_1)$



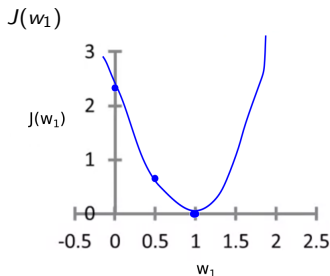
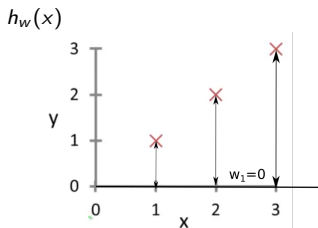
$$J(0.5) = 1/6 * [(0.5 - 1)^2 + (1 - 2)^2 + (1.5 - 3)^2] = 3.5/6 \approx 0.58$$

$$J(0) = 1/6 * [1^2 + 2^2 + 3^2] \approx 2.3$$

Linear regression

- $\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$
- cost function: $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$, thus $\min_{w_0, w_1} J(w_0, w_1)$

Intuition: simplify the hypothesis: $h_w(x) = w_1 x_1$, we want $\min_{w_1} J(w_1)$



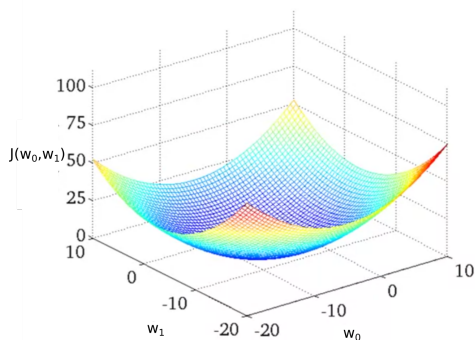
$$J(0.5) = 1/6 * [(0.5 - 1)^2 + (1 - 2)^2 + (1.5 - 3)^2] = 3.5/6 \approx 0.58$$

$$J(0) = 1/6 * [1^2 + 2^2 + 3^2] \approx 2.3$$

Linear regression

Cost function when we do not simplify the hypothesis, i. e.,

$$h_w(x) = w_0 + w_1x$$



Linear regression: gradient descent

Linear regression model:

- $h_w(x) = w_0 + w_1x_1$
- $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$

$$\frac{\partial}{\partial w_j} J(w_0, w_1) = \frac{\partial}{\partial w_j} \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2 = \frac{\partial}{\partial w_j} \frac{1}{2m} \sum_{i=1}^m (w_0 + w_1x^{(i)} - y^{(i)})^2$$

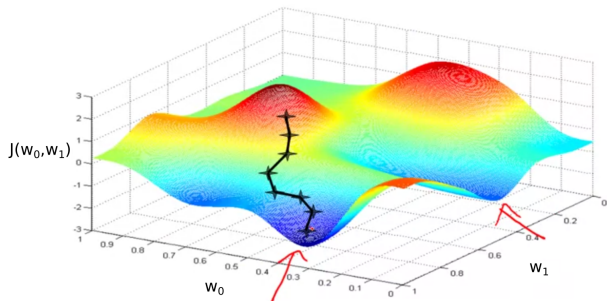
$$\frac{\partial}{\partial w_0} J(w_0, w_1) = \frac{\partial}{\partial w_0} \frac{1}{m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})$$

$$\frac{\partial}{\partial w_1} J(w_0, w_1) = \frac{\partial}{\partial w_1} \frac{1}{m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)}) x^{(i)}$$

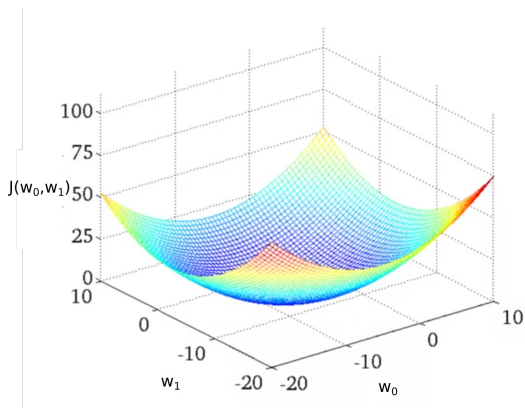
Gradient descent:

repeat until convergence {
 $w_j \leftarrow w_j - \alpha \frac{\partial}{\partial w_j} J(w_0, w_1)$
 for $j = 0$ and $j = 1$
}

Linear regression: gradient descent



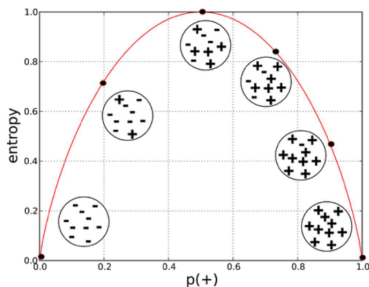
Linear regression: gradient descent



Decision trees

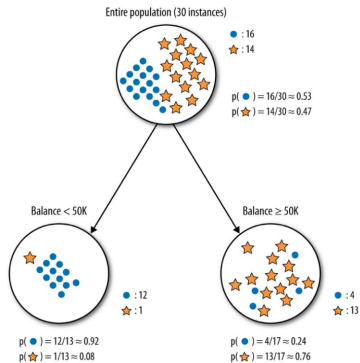
- 1 Select the most important attribute based on entropy $H(V)$
- 2 Divide samples based on this attribute
- 3 Repeat 1) and 2) until all samples belong to the same category

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$



Exercise: Decision tree (1/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

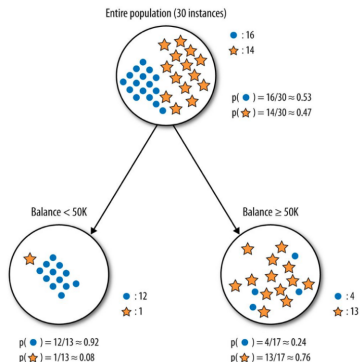


- $H(\text{parent}) =$
- $H(\text{balance} < 50) =$
- $H(\text{balance} \geq 50) =$

Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

Exercise: Decision tree (1/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

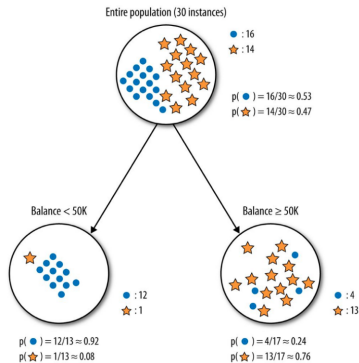


- $H(\text{parent}) = - \frac{16}{30} \log \frac{16}{30} - \frac{14}{30} \log \frac{14}{30} \approx 0.99$
- $H(\text{balance} < 50) =$
- $H(\text{balance} > 50) =$

Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

Exercise: Decision tree (1/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

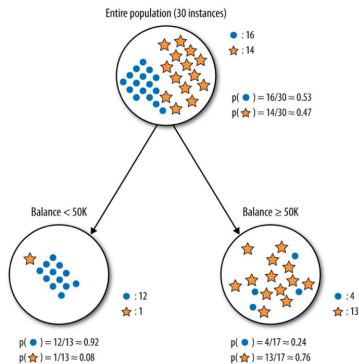


- $H(\text{parent}) = - \frac{16}{30} \log \frac{16}{30} - \frac{14}{30} \log \frac{14}{30} \approx 0.99$
- $H(\text{balance} < 50) = - \frac{12}{13} \log \frac{12}{13} - \frac{1}{13} \log \frac{1}{13} \approx 0.39$
- $H(\text{balance} > 50) =$

Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

Exercise: Decision tree (1/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

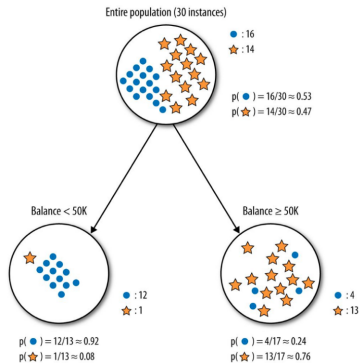


- $H(\text{parent}) = - \frac{16}{30} \log \frac{16}{30} - \frac{14}{30} \log \frac{14}{30} \approx 0.99$
- $H(\text{balance} < 50) = - \frac{12}{13} \log \frac{12}{13} - \frac{1}{13} \log \frac{1}{13} \approx 0.39$
- $H(\text{balance} > 50) = - \frac{4}{17} \log \frac{4}{17} - \frac{13}{17} \log \frac{13}{17} \approx 0.79$

Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

Exercise: Decision tree (1/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

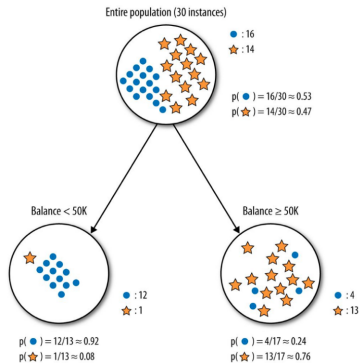


- $H(\text{parent}) = - \frac{16}{30} \log \frac{16}{30} - \frac{14}{30} \log \frac{14}{30} \approx 0.99$
- $H(\text{balance} < 50) = - \frac{12}{13} \log \frac{12}{13} - \frac{1}{13} \log \frac{1}{13} \approx 0.39$
- $H(\text{balance} > 50) = - \frac{4}{17} \log \frac{4}{17} - \frac{13}{17} \log \frac{13}{17} \approx 0.79$

Weighted average of entropy for each node:

Exercise: Decision tree (1/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$



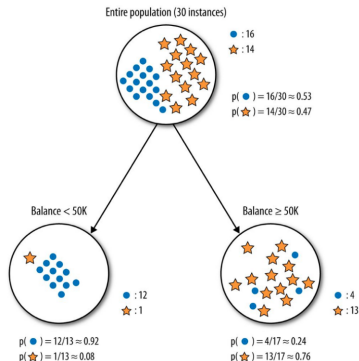
- $H(\text{parent}) = - \frac{16}{30} \log \frac{16}{30} - \frac{14}{30} \log \frac{14}{30} \approx 0.99$
- $H(\text{balance} < 50) = - \frac{12}{13} \log \frac{12}{13} - \frac{1}{13} \log \frac{1}{13} \approx 0.39$
- $H(\text{balance} > 50) = - \frac{4}{17} \log \frac{4}{17} - \frac{13}{17} \log \frac{13}{17} \approx 0.79$

Weighted average of entropy for each node:

- $H(\text{balance}) = 13/30 * 0.39 + 17/30 * 0.79 \approx 0.62$

Exercise: Decision tree (1/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$



- $H(\text{parent}) = - \frac{16}{30} \log \frac{16}{30} - \frac{14}{30} \log \frac{14}{30} \approx 0.99$
- $H(\text{balance} < 50) = - \frac{12}{13} \log \frac{12}{13} - \frac{1}{13} \log \frac{1}{13} \approx 0.39$
- $H(\text{balance} > 50) = - \frac{4}{17} \log \frac{4}{17} - \frac{13}{17} \log \frac{13}{17} \approx 0.79$

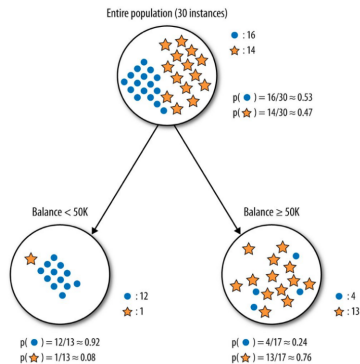
Weighted average of entropy for each node:

- $H(\text{balance}) = 13/30 * 0.39 + 17/30 * 0.79 \approx 0.62$

Information gain:

Exercise: Decision tree (1/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$



- $H(\text{parent}) = -\frac{16}{30} \log \frac{16}{30} - \frac{14}{30} \log \frac{14}{30} \approx 0.99$
- $H(\text{balance} < 50) = -\frac{12}{13} \log \frac{12}{13} - \frac{1}{13} \log \frac{1}{13} \approx 0.39$
- $H(\text{balance} > 50) = -\frac{4}{17} \log \frac{4}{17} - \frac{13}{17} \log \frac{13}{17} \approx 0.79$

Weighted average of entropy for each node:

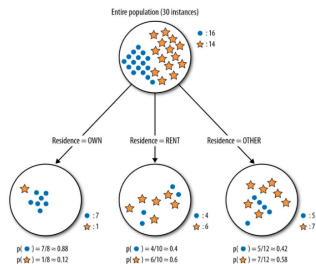
- $H(\text{balance}) = \frac{13}{30} * 0.39 + \frac{17}{30} * 0.79 \approx 0.62$

Information gain:

- $\text{Gain}(\text{parent}, \text{balance}) = H(\text{parent}) - H(\text{balance}) = 0.99 - 0.62 = 0.37$

Exercise: Decision tree (2/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

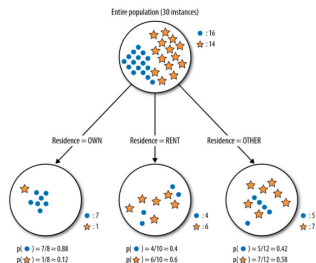


Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

Exercise: Decision tree (2/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

- $H(\text{residence} = \text{OWN}) = -\frac{7}{8} \log \frac{7}{8} - \frac{1}{8} \log \frac{1}{8} \approx 0.54$

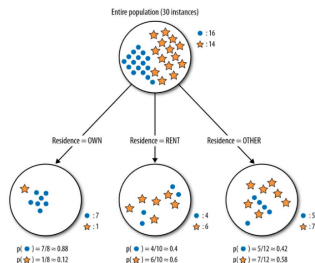


Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

Exercise: Decision tree (2/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

- $H(\text{residence} = \text{OWN}) = -\frac{7}{8} \log \frac{7}{8} - \frac{1}{8} \log \frac{1}{8} \approx 0.54$
- $H(\text{residence} = \text{RENT}) = -\frac{4}{10} \log \frac{4}{10} - \frac{6}{10} \log \frac{6}{10} \approx 0.97$

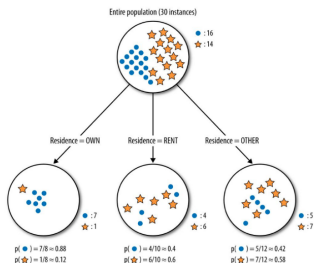


Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

Exercise: Decision tree (2/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

- $H(\text{residence} = \text{OWN}) = -\frac{7}{8} \log \frac{7}{8} - \frac{1}{8} \log \frac{1}{8} \approx 0.54$
- $H(\text{residence} = \text{RENT}) = -\frac{4}{10} \log \frac{4}{10} - \frac{6}{10} \log \frac{6}{10} \approx 0.97$
- $H(\text{residence} = \text{OTHER}) = -\frac{5}{12} \log \frac{5}{12} - \frac{7}{12} \log \frac{7}{12} \approx 0.98$



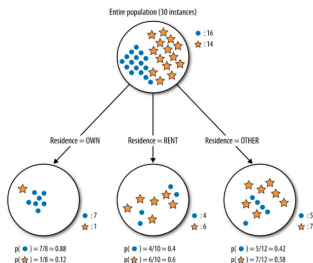
Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

Exercise: Decision tree (2/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

- $H(\text{residence} = \text{OWN}) = -\frac{7}{8} \log \frac{7}{8} - \frac{1}{8} \log \frac{1}{8} \approx 0.54$
- $H(\text{residence} = \text{RENT}) = -\frac{4}{10} \log \frac{4}{10} - \frac{6}{10} \log \frac{6}{10} \approx 0.97$
- $H(\text{residence} = \text{OTHER}) = -\frac{5}{12} \log \frac{5}{12} - \frac{7}{12} \log \frac{7}{12} \approx 0.98$

Weighted average of entropy for each node:



Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

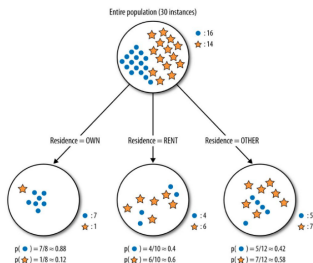
Exercise: Decision tree (2/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

- $H(\text{residence} = \text{OWN}) = -\frac{7}{8} \log \frac{7}{8} - \frac{1}{8} \log \frac{1}{8} \approx 0.54$
- $H(\text{residence} = \text{RENT}) = -\frac{4}{10} \log \frac{4}{10} - \frac{6}{10} \log \frac{6}{10} \approx 0.97$
- $H(\text{residence} = \text{OTHER}) = -\frac{5}{12} \log \frac{5}{12} - \frac{7}{12} \log \frac{7}{12} \approx 0.98$

Weighted average of entropy for each node:

- $H(\text{residence}) = 8/30 * 0.54 + 10/30 * 0.97 + 12/30 * 0.98 \approx 0.86$



Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

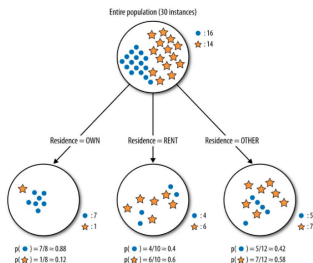
Exercise: Decision tree (2/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$

- $H(\text{residence} = \text{OWN}) = -\frac{7}{8} \log \frac{7}{8} - \frac{1}{8} \log \frac{1}{8} \approx 0.54$
- $H(\text{residence} = \text{RENT}) = -\frac{4}{10} \log \frac{4}{10} - \frac{6}{10} \log \frac{6}{10} \approx 0.97$
- $H(\text{residence} = \text{OTHER}) = -\frac{5}{12} \log \frac{5}{12} - \frac{7}{12} \log \frac{7}{12} \approx 0.98$

Weighted average of entropy for each node:

- $H(\text{residence}) = 8/30 * 0.54 + 10/30 * 0.97 + 12/30 * 0.98 \approx 0.86$

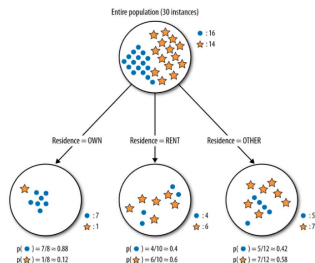


Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

Information gain:

Exercise: Decision tree (2/2)

$$H(V) = - \sum_{i=1}^c p_i \log_2 p_i$$



Provost, Foster; Fawcett, Tom. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

- $H(\text{residence} = \text{OWN}) = -\frac{7}{8} \log \frac{7}{8} - \frac{1}{8} \log \frac{1}{8} \approx 0.54$
- $H(\text{residence} = \text{RENT}) = -\frac{4}{10} \log \frac{4}{10} - \frac{6}{10} \log \frac{6}{10} \approx 0.97$
- $H(\text{residence} = \text{OTHER}) = -\frac{5}{12} \log \frac{5}{12} - \frac{7}{12} \log \frac{7}{12} \approx 0.98$

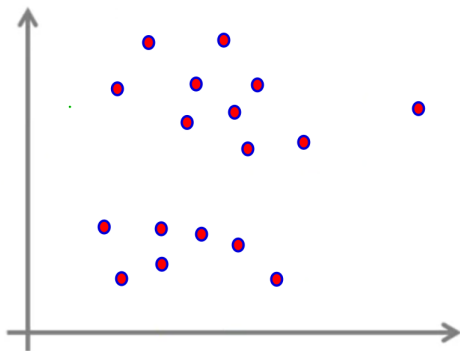
Weighted average of entropy for each node:

- $H(\text{residence}) = 8/30 * 0.54 + 10/30 * 0.97 + 12/30 * 0.98 \approx 0.86$

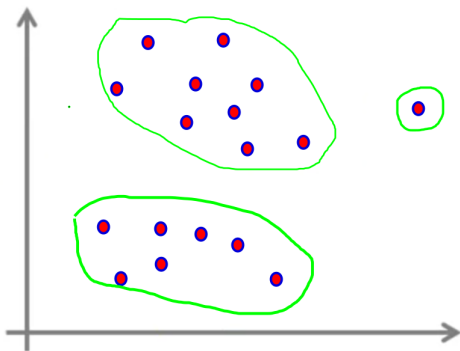
Information gain:

- $\text{Gain}(\text{parent}, \text{residence}) = H(\text{parent}) - H(\text{residence}) = 0.99 - 0.86 = 0.13$

Unsupervised learning: clustering



Unsupervised learning: clustering



Unsupervised learning: clustering example

Google news

The screenshot shows the Google News homepage. The browser address bar displays the URL: <https://news.google.com/topics/CAAqJggKIIBDQkFTRWdwSUwyMHZNRGx1Y1Y4U0FtVnVH20pWVXlnQVAB7hl=en-US&gl=US&ceid=US%3Aen>. The page features a search bar at the top with the text "Search for topics, locations & sources". On the left, there is a navigation menu with categories: Top stories, For you, Following, Saved searches, COVID-19, U.S., World (highlighted), Your local news, Business, Technology, Entertainment, Sports, Science, and Health. Below the menu, there are options for "Language & region" (English (United States)) and "Settings".

The main content area is titled "World" and includes "Follow" and "Share" buttons. It displays a list of news articles:

- Israel-Gaza: Fears of war as violence escalates**
BBC News · 1 hour ago
- Hamas has launched rockets at Tel Aviv after a major Israeli airstrike in Gaza | DW News**
DW News · 14 hours ago
- At least 35 killed in Gaza as Israel ramps up airstrikes in response to rocket attacks**
CNN · 3 hours ago
- Is Iron Dome era dominance over? - analysis**
The Jerusalem Post · 3 hours ago · Local coverage
- Opinion | Why So Much Rests on the Fate of a Tiny Neighborhood in East Jerusalem**
The New York Times · 41 minutes ago · Opinion

A "View Full Coverage" link is visible below the first article. The second article is partially visible:

- Expert panel says mistakes led to coronavirus pandemic, but stops short of holding countries, leaders to account**
The Washington Post · 1 hour ago
- WCNC Charlotte. Always On. Streaming News for May 12, 2021**
WCNC · 2 hours ago
- Covid pandemic was preventable, says WHO-commissioned report**
The Guardian · 48 minutes ago
- Panel suggests WHO should have more power to stop pandemics**
Fox News · 51 minutes ago
- WHO and global leaders could have averted Covid calamity, experts say**

k-means

Simple, widely used clustering algorithm

The number of clusters must be given

Input: k , set of points a_1, \dots, a_n

Place centroids c_1, \dots, c_k at random locations

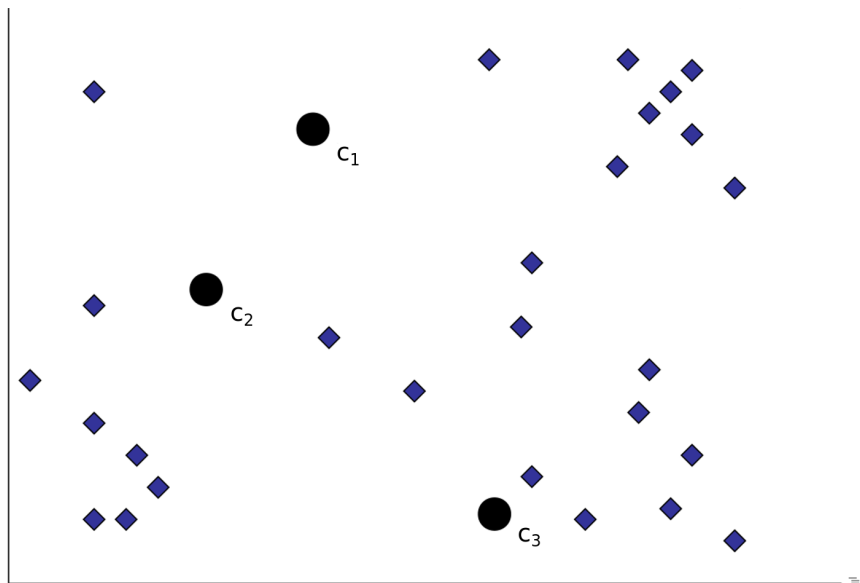
Repeat until convergence:

- for each point a_i :
 - find nearest centroid: $c_j = \arg \min_j d_{ij}$
 - assign the point a_i to cluster C_j
- for each cluster $C_j, j = 1, \dots, k$:
 - new centroid $c_j =$ mean of all points a_i assigned to cluster C_j in the previous step

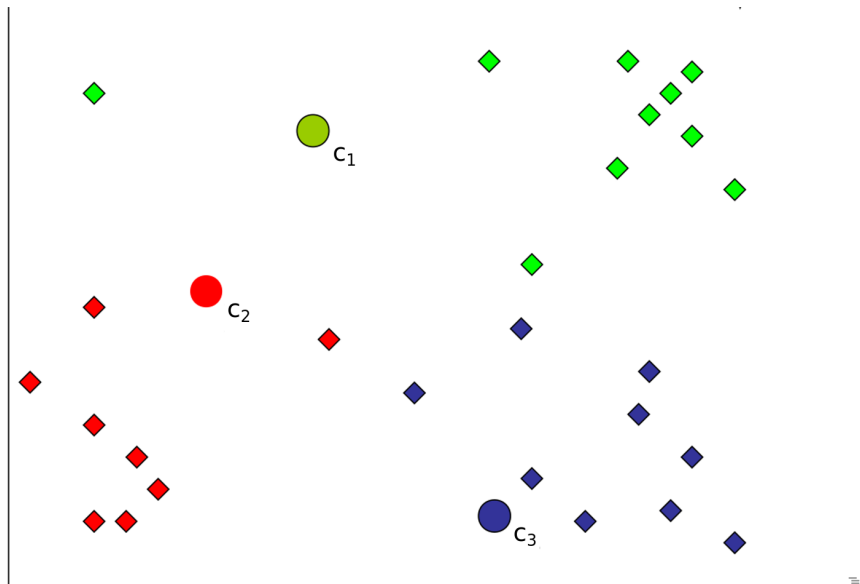
$$c_j(\ell) = \frac{1}{n} \sum_{a_i \in C_j} a_i(\ell) \text{ for } \ell = 1, \dots, m$$

Stop when none of the cluster assignments change

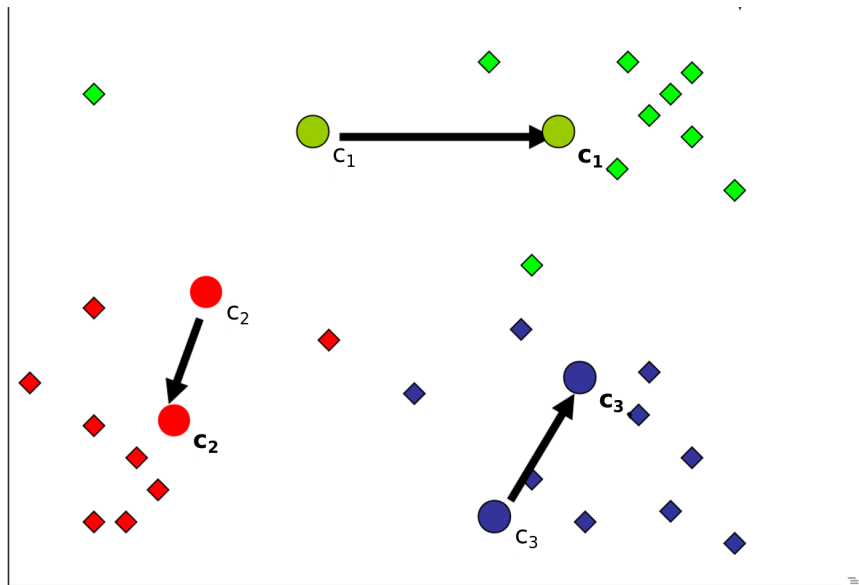
k-means



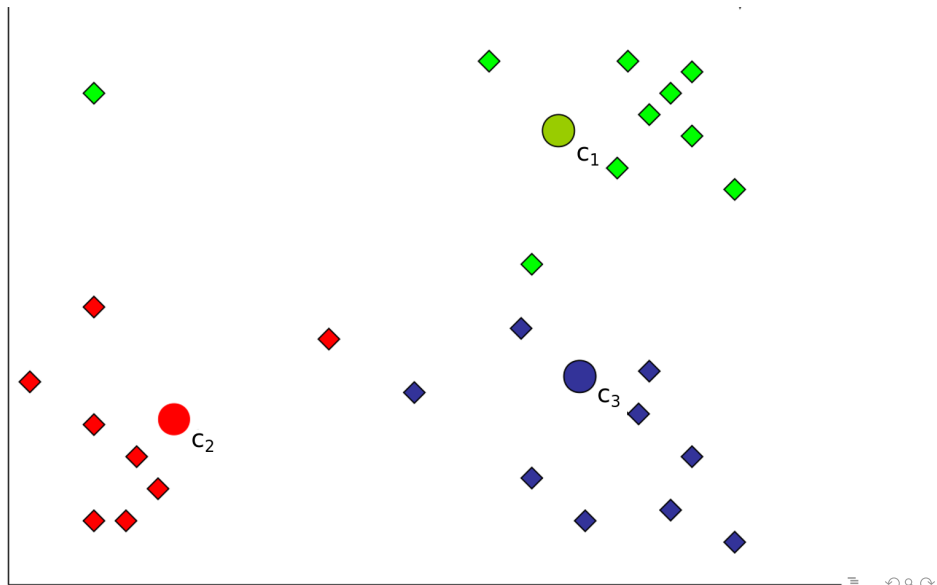
k-means



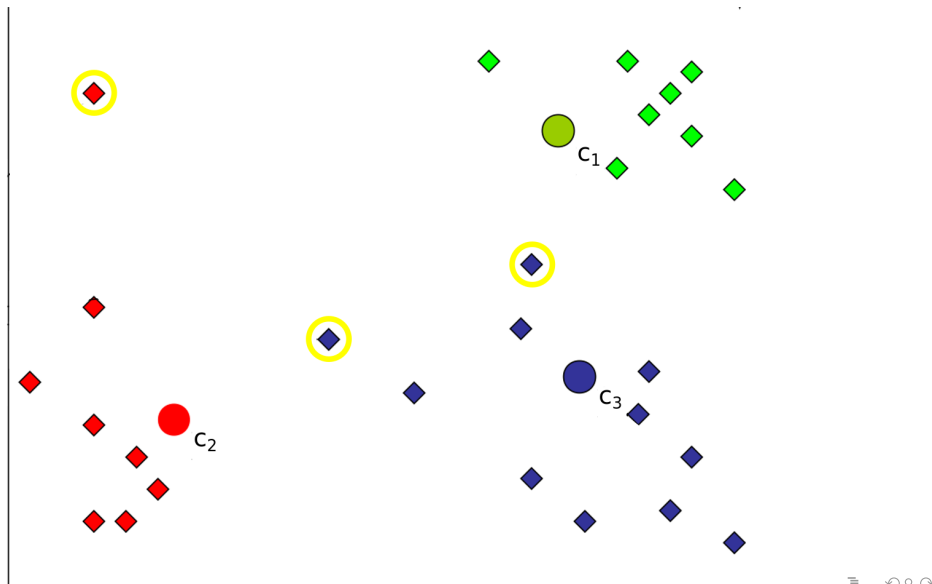
k-means



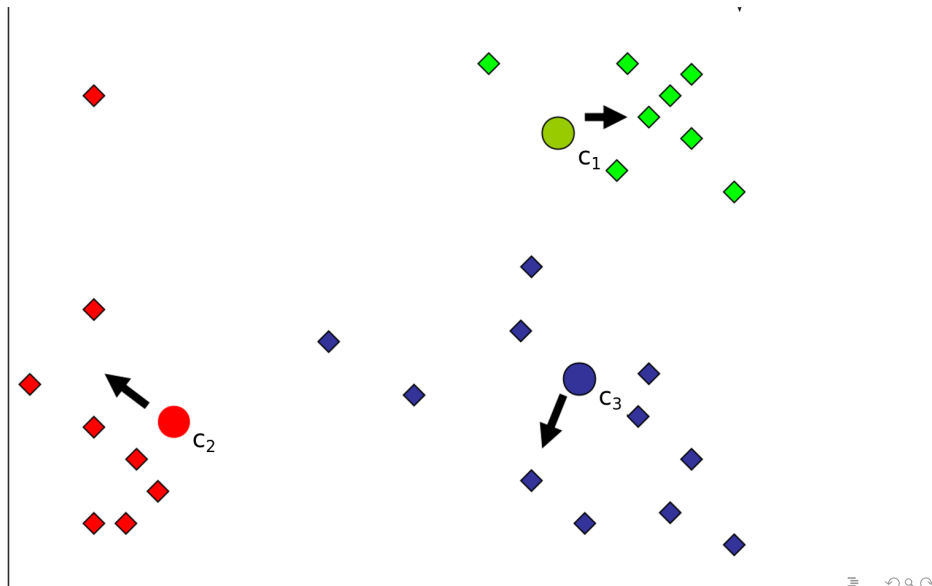
k-means



k -means



k-means



k-means

