Exercise: Examples of AI

Find the most significant and profound example of (one of) the following topics

A. Superhuman AI Artificial intelligence that outperforms humans

`https://finnaarupnielsen.wordpress.com/2015/03/15/status-on-human-vs-machines`

B. Emulating human creativity AI that emulates human creativity

`http://www.thepaintingfool.com`

C. Intelligent animal behavior Animal behavior

`https://www.thespruce.com/understanding-bird-intelligence-386440`,

`https://en.wikipedia.org/wiki/Dog_intelligence`

D. Augmented intelligence Enhancing human performance using AI

`https://www.technologyreview.com/s/603951/this-is-your-brain-on-gps-navigation`,

`https://deepmind.com/blog/2017-deepminds-year-review`

Prepare to present your example with *two sentences*:

1. Describe the example briefly.
2. Describe why you think this example is significant.

Exercise: Turings objections

Is it possible to create a *thinking machine*? Turing outlined 9 objections:

Theological	Only God can create thinking machines.
Heads in the sand	The consequences of thinking machines are too dreadful.
Mathematical	Fundamental limitations to the power of state machines.
Consciousness	The machine can merely imitate—it cannot feel.
Disabilities	Okay you can do all these things, but you can't do X...
Determinism	The machine can only do what we tell it.
Discrete	The human nervous system is continuous.
Informality	We cannot define rules for every conceivable circumstance.
Extra-sensory	As-of-yet undiscovered laws of physics govern thinking.

Discuss in groups

- Do you belive it is possible to create a thinking machine?
- Which of these objections you agree/disagree with
- Can you come up with any other objections?

Prepare to present your argument for or against thinking machines.

Exercise: Population mean and variance

Consider a population of $N = 3$ observations.

$$x = \{1, 4, 10\}$$

- What is the population mean μ_x and variance σ_x^2?

Definitions

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2$$

Exercise: Population mean and variance

Consider a population of $N = 3$ observations.

$$x = \{1, 4, 10\}$$

- What is the population mean μ_x and variance σ_x^2?

Definitions

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2$$

Solution

$$\mu_x = \frac{1}{3}(1 + 4 + 10) = 5$$

$$\sigma_x^2 = \frac{1}{3}\left((1-5)^2 + (4-5)^2 + (10-5)^2\right) = 14$$

Exercise: Why divide by $n-1$?

Consider a population of $N = 3$ observations

$$x = \{1, 4, 10\}$$

with population mean and variance

$$\mu_x = 5 \qquad \sigma_x^2 = 14$$

- List all possible ordered samples with replacement of size $n = 2$. (Hint: There are 9 such possible samples)

Exercise: Why divide by $n-1$?

Consider a population of $N = 3$ observations

$$x = \{1, 4, 10\}$$

with population mean and variance

$$\mu_x = 5 \qquad \sigma_x^2 = 14$$

- List all possible ordered samples with replacement of size $n = 2$. (Hint: There are 9 such possible samples)

Solution
The 9 possible samples are

$$\{1,1\},\{1,4\},\{1,10\},\{4,1\},\{4,4\},\{4,10\},\{10,1\},\{10,4\},\{10,10\}$$

Exercise: Why divide by $n-1$? (II)

Consider a population of $N = 3$ observations

$$x = \{1, 4, 10\}$$

with population mean and variance

$$\mu_x = 5 \qquad \sigma_x^2 = 14$$

Sample estimate

$$m_x = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_x)^2$$

- Compute the sample estimate of the mean and variance, m_x and s_{n-1}^2 for each possible sample

$$\{1,1\},\{1,4\},\{1,10\},\{4,1\},\{4,4\},\{4,10\},\{10,1\},\{10,4\},\{10,10\}$$

- What is the average sample estimate of the mean and variance (averaged over all possible samples)?

Exercise: Why divide by $n-1$? (II)

Solution

Sample	m_x	s_x^2
$\{1,1\}$		
$\{1,4\}$		
$\{1,10\}$		
$\{4,1\}$		
$\{4,4\}$		
$\{4,10\}$		
$\{10,1\}$		
$\{10,4\}$		
$\{10,10\}$		

Exercise: Why divide by $n-1$? (II)

Solution

Sample	m_x	s_x^2
$\{1,1\}$	$\frac{1+1}{2}=1$	$\frac{(1-1)^2+(1-1)^2}{2-1}=0$
$\{1,4\}$		
$\{1,10\}$		
$\{4,1\}$		
$\{4,4\}$		
$\{4,10\}$		
$\{10,1\}$		
$\{10,4\}$		
$\{10,10\}$		

Exercise: Why divide by $n-1$? (II)

Solution

Sample	m_x	s_x^2
$\{1,1\}$	$\frac{1+1}{2}=1$	$\frac{(1-1)^2+(1-1)^2}{2-1}=0$
$\{1,4\}$	$\frac{1+4}{2}=2.5$	$\frac{(1-2.5)^2+(4-2.5)^2}{2-1}=4.5$
$\{1,10\}$		
$\{4,1\}$		
$\{4,4\}$		
$\{4,10\}$		
$\{10,1\}$		
$\{10,4\}$		
$\{10,10\}$		

Exercise: Why divide by $n-1$? (II)

Solution

Sample	m_x	s_x^2
$\{1,1\}$	$\frac{1+1}{2}=1$	$\frac{(1-1)^2+(1-1)^2}{2-1}=0$
$\{1,4\}$	$\frac{1+4}{2}=2.5$	$\frac{(1-2.5)^2+(4-2.5)^2}{2-1}=4.5$
$\{1,10\}$	$\frac{1+10}{2}=5.5$	$\frac{(1-5.5)^2+(10-5.5)^2}{2-1}=40.5$
$\{4,1\}$		
$\{4,4\}$		
$\{4,10\}$		
$\{10,1\}$		
$\{10,4\}$		
$\{10,10\}$		

Exercise: Why divide by $n-1$? (II)

Solution

Sample	m_x	s_x^2
$\{1,1\}$	$\frac{1+1}{2}=1$	$\frac{(1-1)^2+(1-1)^2}{2-1}=0$
$\{1,4\}$	$\frac{1+4}{2}=2.5$	$\frac{(1-2.5)^2+(4-2.5)^2}{2-1}=4.5$
$\{1,10\}$	$\frac{1+10}{2}=5.5$	$\frac{(1-5.5)^2+(10-5.5)^2}{2-1}=40.5$
$\{4,1\}$	2.5	4.5
$\{4,4\}$	4	0
$\{4,10\}$	7	18
$\{10,1\}$	5.5	40.5
$\{10,4\}$	7	18
$\{10,10\}$	10	0

Exercise: Why divide by $n - 1$? (II)

Solution

Sample	m_x	s_x^2
$\{1,1\}$	$\frac{1+1}{2} = 1$	$\frac{(1-1)^2+(1-1)^2}{2-1} = 0$
$\{1,4\}$	$\frac{1+4}{2} = 2.5$	$\frac{(1-2.5)^2+(4-2.5)^2}{2-1} = 4.5$
$\{1,10\}$	$\frac{1+10}{2} = 5.5$	$\frac{(1-5.5)^2+(10-5.5)^2}{2-1} = 40.5$
$\{4,1\}$	2.5	4.5
$\{4,4\}$	4	0
$\{4,10\}$	7	18
$\{10,1\}$	5.5	40.5
$\{10,4\}$	7	18
$\{10,10\}$	10	0

Average s_x^2 over all possible samples

$$\operatorname{avg}(m_x) = \frac{1 + 2.5 + 5.5 + 2.5 + 4 + 7 + 5.5 + 7 + 10}{9} = \frac{45}{9} = 5 = \mu_x$$

$$\operatorname{avg}(s_x^2) = \frac{0 + 4.5 + 40.5 + 4.5 + 0 + 18 + 40.5 + 18 + 0}{9} = \frac{126}{9} = 14 = \sigma_x^2$$

Exercise: Mean and variance of a 6-sided dice

Mean and standard deviation of a discrete distribution

- Sum over all possible outcomes
- Weigh each by their probability

$$\mu_x = \sum_{k=1}^{K} P(x_k) \cdot x_k \qquad \sigma_x^2 = \sum_{k=1}^{K} P(x_k) \cdot (x_k - \mu)^2$$

- What is μ_x and σ_x^2 for a normal 6-sided dice?
  $K = 6, \quad x_1 = 1, x_2 = 2, \ldots, x_6 = 6, \quad P(x_1) = P(x_2) = \cdots = P(x_6) = \frac{1}{6}$

Exercise: Mean and variance of a 6-sided dice

Mean and standard deviation of a discrete distribution

- Sum over all possible outcomes
- Weigh each by their probability

$$\mu_x = \sum_{k=1}^{K} P(x_k) \cdot x_k \qquad \sigma_x^2 = \sum_{k=1}^{K} P(x_k) \cdot (x_k - \mu)^2$$

- What is μ_x and σ_x^2 for a normal 6-sided dice?
  $K = 6, \quad x_1 = 1, x_2 = 2, \ldots, x_6 = 6, \quad P(x_1) = P(x_2) = \cdots = P(x_6) = \frac{1}{6}$

Solution

$$\mu_x = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{21}{6} = 3.5$$

$$\sigma_x^2 = \frac{1}{6} \left((1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2 \right)$$

$$= \frac{1}{6}(6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25) \approx 2.917$$

Exercise: Confidence interval of 10 dice throws

- Throw a 6-side dice 10 times and record the results
  (e.g. use `www.random.org/dice`)
- Compute the 50% confidence interval for the mean
  Express it as a range [low, high]

Confidence interval

$$m_x \pm \underbrace{z_{\alpha/2}}_{\text{critical value}} \cdot \underbrace{\sqrt{\frac{s_x^2}{n}}}_{\text{standard error}}$$

$(z_{0.25} = 0.67)$

Exercise: Confidence interval of 10 dice throws

- Throw a 6-side dice 10 times and record the results
  (e.g. use `www.random.org/dice`)
- Compute the 50% confidence interval for the mean
  Express it as a range [low, high]

Confidence interval

$$m_x \pm \underbrace{z_{\alpha/2}}_{\text{critical value}} \cdot \underbrace{\sqrt{\frac{s_x^2}{n}}}_{\text{standard error}}$$

$$(z_{0.25} = 0.67)$$

Solution example

Exercise: Confidence interval of 10 dice throws

- Throw a 6-side dice 10 times and record the results
  (e.g. use `www.random.org/dice`)
- Compute the 50% confidence interval for the mean
  Express it as a range [low, high]

Confidence interval

$$m_x \pm \underbrace{z_{\alpha/2}}_{\text{critical value}} \cdot \underbrace{\sqrt{\frac{s_x^2}{n}}}_{\text{standard error}}$$

$$(z_{0.25} = 0.67)$$

Solution example

$$m_x = \frac{1}{10}(1+6+5+6+1+6+3+1+3+1) = \frac{33}{10} = 3.3$$

$$s_x^2 = \frac{1}{10-1}\left((1-3.3)^2 + (6-3.3)^2(5-3.3)^2 + + \cdots + (1-1)^2\right) \approx 5.12$$

Confidence interval

$$m_x \pm z_{\alpha/2} \cdot \sqrt{\frac{s_x^2}{n}} = 3.3 \pm 0.67 \cdot \sqrt{\frac{5.12}{10}} = 3.3 \pm 0.48$$

$$[2.82, 3.78]$$

The population mean is 3.5 and we expect 50% of the computed confidence intervals to include it

Exercise: An algorithm for sorting

1. Write down the numbers below on 8 small pieces of paper
2. Lay them in a random sequence on the table
3. Sort them starting with the smallest, and take notice of exactly which procedure you use
4. Write down a high level description of your sorting algorithm
5. Randomize the order of the numbers again, and follow your written procedure to the letter to sort the numbers again

Example list

99 83 125 12 5 256 31 192

Prepare to present your algorithm to the class

Exercise: Merge sort

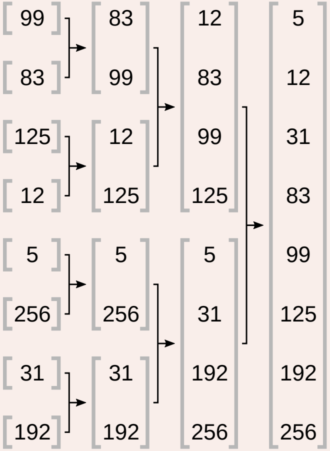
Algorithm

At all times, maintain a set of sorted sublists
Initially each element is a sorted sublist

1. Merge each pair of sublists to form a new sorted sublist
2. Repeat until all sublists have been merged

Question

- How many operations (comparisons) are required (in the worst case) to sort a list of 8 items?
- What is the algorithmic complexity of merge sort?
  Assume for simplicity that the number of elements is a power of two, $n = 2^\ell$.

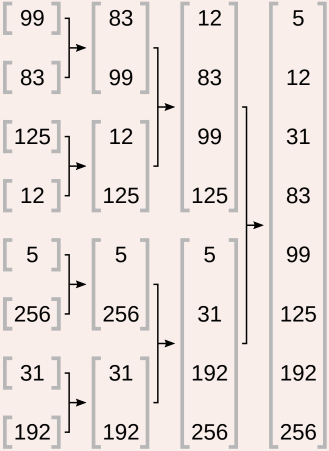


Exercise: Merge sort

Algorithm

At all times, maintain a set of sorted sublists
Initially each element is a sorted sublist

1. Merge each pair of sublists to form a new sorted sublist
2. Repeat until all sublists have been merged



Question

- How many operations (comparisons) are required (in the worst case) to sort a list of 8 items?
- What is the algorithmic complexity of merge sort?
  Assume for simplicity that the number of elements is a power of two, $n = 2^\ell$.

Solution

$$T(n) = \frac{n}{2}\cdot 1 + \frac{n}{4}\cdot 3 + \frac{n}{8}\cdot 7 + \cdots = n(\frac{1}{2} + \frac{3}{4} + \frac{7}{8} + \cdots) < n\ell \quad T(8) = 17$$

$$n = 2^\ell \Leftrightarrow \ell = \log_2(n), \quad T(n) \in O(n\log n)$$

Exercise: Dot product

We can use the dot product between the word occurence vectors as a measure of similarity between documents

- Compute the dot product between the two sentences
- Can you think of pros and cons of using the dot product to measure similarity?

Sentences

1. `Zebras are several species of African equids (horse family) united by their distinctive black and white striped coats.`
2. `Although the okapi bears striped markings reminiscent of zebras it is most closely related to the giraffe.`

Words in common in the sentences

	Doc. 1	Doc. 2
`of`	1	1
`stripe`	1	1
`zebra`	1	1

Exercise: TF-IDF

- Consider a document that contains 100 words, wherein
  - the word `the` appears 3 times and
  - the word `cat` appears 3 times
- The document is part of a 10 000 document corpus, wherein
  - 4 900 of the documents contain the word `the` and
  - 123 of the documents contain the word `cat`

Compute the TF and IDF for the terms `the` and `cat`

TF and IDF

$$TF = \frac{n_{t,d}}{n_d} \qquad IDF = \log\left(\frac{N}{n_t}\right)$$

- $n_{t,d}$ Number of occurences of term t in document d
- n_d Number of terms in document d
- n_t Number of documents with term t
- N Total number of documents

Exercise: TF-IDF

- Consider a document that contains 100 words, wherein
  - the word *the* appears 3 times and
  - the word *cat* appears 3 times
- The document is part of a 10 000 document corpus, wherein
  - 4 900 of the documents contain the word *the* and
  - 123 of the documents contain the word *cat*

Compute the TF and IDF for the terms *the* and *cat*

TF and IDF

$$TF = \frac{n_{t,d}}{n_d} \qquad IDF = \log\left(\frac{N}{n_t}\right)$$

- $n_{t,d}$ Number of occurences of term t in document d
- n_d Number of terms in document d
- n_t Number of documents with term t
- N Total number of documents

Solution

the

$$\text{TF} = \frac{3}{100} = 0.03$$

$$\text{IDF} = \log\left(\frac{10\,000}{4\,900}\right) \approx 0.7133$$

cat

$$\text{TF} = \frac{3}{100} = 0.03$$

$$\text{IDF} = \log\left(\frac{10\,000}{123}\right) \approx 4.398$$

Exercise: TF-IDF

- What happens if no documents contain one of the search terms?

TF-IDF

$$\text{TF-IDF}(d, q) = \sum_{t \in q} \frac{n_{t,d}}{n_d} \cdot \log\left(\frac{N}{n_t}\right)$$

$n_{t,d}$ Number of occurences of term t in document d

n_d Number of terms in document d

n_t Number of documents with term t

N Total number of documents

Exercise: TF-IDF

- What happens if no documents contain one of the search terms?

TF-IDF

$$\text{TF-IDF}(d, q) = \sum_{t \in q} \frac{n_{t,d}}{n_d} \cdot \log\left(\frac{N}{n_t}\right)$$

$n_{t,d}$ Number of occurences of term t in document d

n_d Number of terms in document d

n_t Number of documents with term t

N Total number of documents

Solution: Division by zero!

Exercise: Okapi BM25

BM25

$$\text{BM25}(d, q) = \sum_{t \in q} \frac{n_{t,d} \cdot (k_1 + 1)}{n_{t,d} + k_1 \cdot (1 - b + b \cdot \frac{n_d}{\text{avgdl}})} \cdot \log\left(\frac{N - n_t + 0.5}{n_t + 0.5}\right)$$

- Consider a document that contains 100 words, wherein
  - the word *the* appears 3 times and
  - the word *cat* appears 3 times
- The document is part of a 10 000 document corpus, wherein
  - 4 900 of the documents contain the word *the* and
  - 123 of the documents contain the word *cat*
- The average document length in the corpus is 150

$n_{t,d}$	Number of occurences of term t in document d
n_d	Number of terms in document d
n_t	Number of documents with term t
N	Total number of documents
avgdl	Average document length
b	$b = 0.75$
k_1	$k_1 = 1.2$

Compute the BM25-score for the query *the cat*

Exercise: Okapi BM25

Solution

$$
\begin{aligned}
\mathrm{BM25}(d,q) &= \sum_{t\in q} \frac{n_{t,d}\cdot(k_1+1)}{n_{t,d}+k_1\cdot(1-b+b\cdot\frac{n_d}{\mathrm{avgdl}})}\cdot\log\left(\frac{N-n_t+0.5}{n_t+0.5}\right)\\
&= \frac{3\cdot(1.2+1)}{3+1.2\cdot(1-0.75+0.75\cdot\frac{100}{150})}\cdot\log\left(\frac{10\,000-4\,900+0.5}{4\,900+0.5}\right)+\\
&\quad \frac{3\cdot(1.2+1)}{3+1.2\cdot(1-0.75+0.75\cdot\frac{100}{150})}\cdot\log\left(\frac{10\,000-123+0.5}{123+0.5}\right)\\
&\approx 1.692\cdot 0.040+1.692\cdot 4.382\approx \underline{7.483}
\end{aligned}
$$

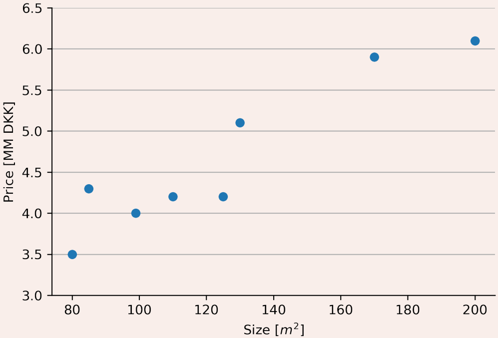
Exercise: What is human learning?

Is human learning best characterized as

- Unsupervised learning
- Supervised learning
- Reinforcement learning

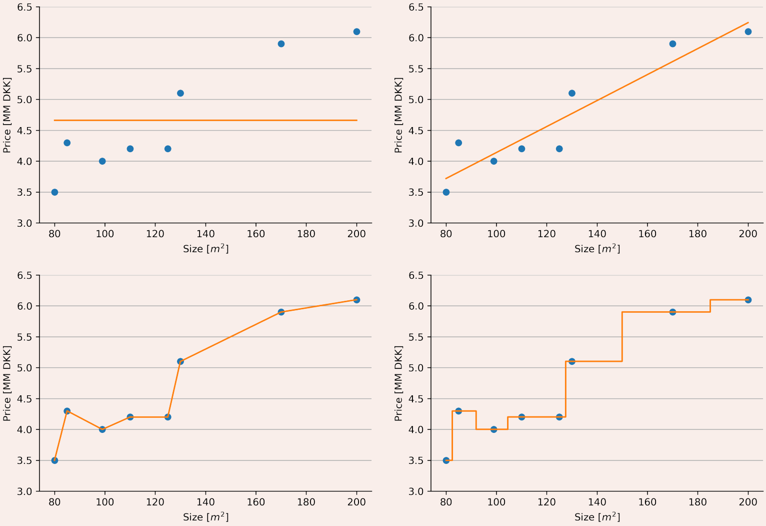
(If you think the answer is somehow obvious, see if you can come up with an argument against)

Exercise: Price of a 150 m^2 house



- What would you expect the price of a 150 m^2 house to be?
- Discuss which “algorithm” you used to come up with your answer

Exercise: House price regression



- Which of the above regression curves is best?
- Discuss how you could define a criteria for which is "best"

Exercise: Least squares regression

Solve the least square regression problem by minimizing the error

- Differentiate the error measure wrt. the parameters a and b
- This gives you two equations in two unknowns to solve

Problem specification

- Data

$$x = \{80, 85, 99, 110, 125, 130, 170, 200\}$$
$$y = \{3.5, 4.3, 4, 4.2, 4.2, 5.1, 5.9, 6.1\}$$

- Regression function

$$f(x) = ax + b$$

- Error measure

$$E = \sum_{n=1}^{N} (y_n - f(x_n))^2$$

Some useful definitions

$$\bar{x} = \sum_{n=1}^{N} x_n = 999$$

$$\bar{y} = \sum_{n=1}^{N} y_n = 37.3$$

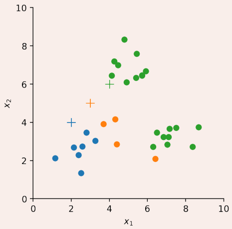
$$\overline{xy} = \sum_{n=1}^{N} x_n y_n = 4914.5$$

$$\overline{xx} = \sum_{n=1}^{N} x_n^2 = 136951$$

Exercise: Optimal cluster center

Fix cluster assignments, optimize cluster means

$$\min_{\{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K\}} \underbrace{\sum_{k=1}^{K}}_{\text{Clusters}} \underbrace{\sum_{n:c_n=k}}_{\substack{\text{Observations}\\ \text{in cluster } k}} \|\boldsymbol{x}_n - \boldsymbol{z}_k\|^2$$

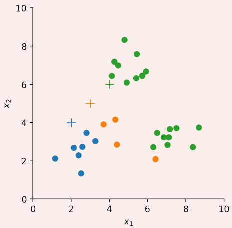


- What is the optimum value of the cluster means $\boldsymbol{z}_k$?
- Hint: Optimize the expression by computing the derivative wrt. $\boldsymbol{z}_k$, equate to zero and solve for $\boldsymbol{z}_k$

Exercise: Optimal cluster center

Fix cluster assignments, optimize cluster means

$$\min_{\{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K\}} \underbrace{\sum_{k=1}^{K}}_{\text{Clusters}} \underbrace{\sum_{n:c_n=k}}_{\substack{\text{Observations}\\ \text{in cluster } k}} \|\boldsymbol{x}_n - \boldsymbol{z}_k\|^2$$



- What is the optimum value of the cluster means $\boldsymbol{z}_k$?
- Hint: Optimize the expression by computing the derivative wrt. $\boldsymbol{z}_k$, equate to zero and solve for $\boldsymbol{z}_k$

Solution

$$\frac{\partial L}{\partial \boldsymbol{z}_k} \sum_{n:c_n=k} -2(\boldsymbol{x}_n - \boldsymbol{z}_k) = 2N_k\boldsymbol{z}_k - 2\sum_{n:c_n=k} \boldsymbol{x}_n = 0 \Rightarrow \boldsymbol{z}_k = \frac{1}{N_k}\sum_{n:c_n=k} \boldsymbol{x}_n$$

Exercise: Pen-and-paper k-means

Using pen-and-paper k-means, cluster the following 1-dimensional data objects

Data	$\{10, 18, 32, 70, 81, 89\}$
Num. clusters	$K = 2$
Initialization	Set means to the first two data points

Algorithm

1. Fix cluster means
   Assign each observation to closest cluster
2. Fix cluster assignments
   Set cluster means to average of data points in cluster

Exercise: K-means computational complexity

- What is the computational complexity of the k-means algorithm?
- Express it in big-O notation in terms of the number of data points N and the number of clusters K

Algorithm

1. Fix cluster means, optimize cluster assignment

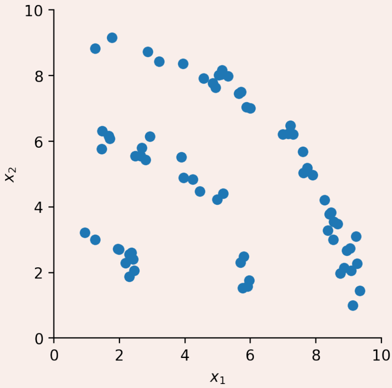
$$\min_{\{c_1,\ldots,c_N\}} \sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{z}_{c_n}\|^2$$

2. Fix cluster assignments, optimize cluster means

$$\min_{\{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K\}} \sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{z}_{c_n}\|^2$$

Exercise: Transformation of input features

- Can you come up with a way to transform the input features, so that k-means will find the three clusters?



Exercise: What is an image?

- Try to make a definition of what an *image* is without using technical terms such as pixels etc.

Exercise: Gradient calculation

Multivariate function

$$f(x, y) = x^2 \cos(y)$$

What is the gradient?

Gradient definition

$$\nabla f(x, y) = \begin{bmatrix} \dfrac{\partial f(x, y)}{\partial x} \\ \\ \dfrac{\partial f(x, y)}{\partial y} \end{bmatrix}$$

Exercise: Gradient calculation

Multivariate function

$$f(x, y) = x^2 \cos(y)$$

What is the gradient?

Partial derivatives

$$\frac{\partial f(x, y)}{\partial x} = 2x\cos(y)$$

$$\frac{\partial f(x, y)}{\partial y} = -x^2 \sin(y)$$

Gradient

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x\cos(y) \\ -x^2 \sin(y) \end{bmatrix}$$

Gradient definition

$$\nabla f(x, y) = \begin{bmatrix} \dfrac{\partial f(x, y)}{\partial x} \\ \\ \dfrac{\partial f(x, y)}{\partial y} \end{bmatrix}$$

Exercise: Gradient of neural network

Compute the partial derivatives

$$\frac{\partial E}{\partial c}, \quad \frac{\partial E}{\partial w_1}, \quad \frac{\partial E}{\partial b_1}, \quad \frac{\partial E}{\partial v_1}$$

Hints

1. Use the chain rule
2. $\dfrac{\partial \tanh(x)}{\partial x} = 1 - \tanh^2(x)$
3. Don't expand terms needlessly. Express in terms of e.g. $\hat{y}_n$ and $h_1(x_n)$ where possible.

Cost function

$$E = \sum_{n=1}^{N} \left(y_n - \hat{y}_n\right)^2$$

Neural network model

$$\begin{aligned} \hat{y}_n &= w_1 h_1(x_n) + w_2 h_2(x_n) + c \\ h_1(x_n) &= \tanh\left(v_1 x_n + b_1\right) \\ h_2(x_n) &= \tanh\left(v_2 x_n + b_2\right) \end{aligned}$$

Exercise: Chain rule

Compute the derivative $\frac{dz}{dt}$ of the following function

$$z(t) = f(x, y) = xy + x^2$$

where

$$x(t) = \sin(t)$$
$$y(t) = t^2$$

Exercise: Chain rule

Compute the derivative $\frac{dz}{dt}$ of the following function

$$z(t) = f(x, y) = xy + x^2$$

where

$$x(t) = \sin(t)$$
$$y(t) = t^2$$

Solution

$$\frac{dz}{dt} = \frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt}$$
$$= (y + 2x) \cdot \cos(t) + x \cdot (2t)$$

Exercise: Computation graph

Draw the computation graph for the function

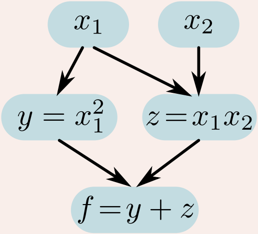
$$f(x_1, x_2) = x_1^2 + x_1 \cdot x_2$$

Exercise: Computation graph

Draw the computation graph for the function

$$f(x_1, x_2) = x_1^2 + x_1 \cdot x_2$$

Solution



Exercise: $f(x) = Ax^2 + Bx + C$

Consider the function

$$f(x) = Ax^2 + Bx + C$$

Evaluate the function on $x = a + \epsilon$

Exercise: $f(x) = Ax^2 + Bx + C$

Consider the function

$$f(x) = Ax^2 + Bx + C$$

Evaluate the function on $x = a + \epsilon$

Solution

$$\begin{aligned} f(a+\epsilon) &= A(a+\epsilon)^2 + B(a+\epsilon) + C \\ &= A(a^2 + 2a\epsilon + \epsilon^2) + B(a+\epsilon) + C \\ &= \underbrace{(Aa^2 + Ba + C)}_{f(a)} + \epsilon \underbrace{(2Aa + B)}_{f'(a)} \end{aligned}$$

Exercise: Audio as a point in a vector space

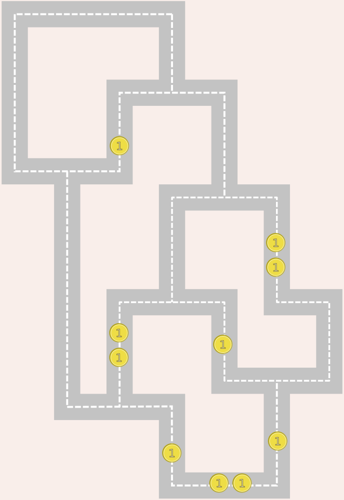
An audio signal of length N can be thought of as a point in an N-dimensional vector space, $\mathbb{R}^N$

- What is the standard basis of this vector space?
- How can we construct any possible audio signal by a linear combination of such basis vectors?
- How do you think each of these basis vectors sounds
- Is this a good basis for representing sound? Can you come up with a better basis, perhaps inspired by the human auditory system?

Exercise: Collecting gold coins

Consider a game, where we drive around and collect gold coins (coins can be picked up multiple times.)
How could we meaningfully define:

- The set of states
- The set of actions
- The next-state function
- The reward



Exercise: Optimal policy

What is the *optimal policy*?
Hint: What should we end up doing, if we follow the optimal policy?