

# COMP9417 - Machine Learning

## Tutorial: Revision

### Question 1. (Calculus Review)

(a) Consider the function

$$f(x, y) = a_1x^2y^2 + a_4xy + a_5x + a_7$$

compute all first and second order derivatives of  $f$  with respect to  $x$  and  $y$ .

**Solution:**

We have the following results:

$$\begin{aligned}\frac{\partial f}{\partial x} &= 2a_1xy^2 + a_4y + a_5 \\ \frac{\partial f}{\partial y} &= 2a_1x^2y + a_4x \\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial^2 f}{\partial y \partial x} = 4a_1xy + a_4 \\ \frac{\partial^2 f}{\partial x^2} &= 2a_1y^2 \\ \frac{\partial^2 f}{\partial y^2} &= 2a_1x^2\end{aligned}$$

(b) Consider the function

$$f(x, y) = a_1x^2y^2 + a_2x^2y + a_3xy^2 + a_4xy + a_5x + a_6y + a_7$$

compute all first and second order derivatives of  $f$  with respect to  $x$  and  $y$ .

**Solution:**

We have the following results:

$$\begin{aligned}\frac{\partial f}{\partial x} &= 2a_1xy^2 + 2a_2xy + a_3y^2 + a_4y + a_5 \\ \frac{\partial f}{\partial y} &= 2a_1x^2y + a_2x^2 + 2a_3xy + a_4x + a_6 \\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial^2 f}{\partial y \partial x} = 4a_1xy + 4a_2x + 2a_3y + a_4 \\ \frac{\partial^2 f}{\partial x^2} &= 2a_1y^2 + 2a_2y \\ \frac{\partial^2 f}{\partial y^2} &= 2a_1x^2 + 2a_3x\end{aligned}$$

(c) Consider the logistic sigmoid:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

show that  $\sigma'(x) = \frac{\partial \sigma}{\partial x} = \sigma(x)(1 - \sigma(x))$

**Solution:**

Rewriting  $\sigma(x) = (1 + e^{-x})^{-1}$ , then using the chain rule gives

$$\begin{aligned}\sigma'(x) &= (-1)(-e^{-x})(1 + e^{-x})^{-2} \\ &= \frac{e^{-x}}{1 + e^{-x}} \frac{1}{1 + e^{-x}} \\ &= (1 - \sigma(x))\sigma(x).\end{aligned}$$

(d) Consider the following functions:

- $y_1 = 4x^2 - 3x + 3$
- $y_2 = 3x^4 - 2x^3$
- $y_3 = 4x + \sqrt{1 - x}$
- $y_4 = x + x^{-1}$

Using the second derivative test, find all local maximum and minimum points.

**Solution:**

For the first function, we differentiate and set equal to zero to find the critical point(s) of the function:

$$y'_1 := \frac{\partial y_1}{\partial x} = 8x - 3 \stackrel{\text{set}}{=} 0 \implies x = \frac{3}{8}.$$

To assess whether  $y_1$  achieves a minimum or maximum at  $x = 3/8$ , we compute the second derivative:

$$y_1'' := \frac{\partial^2 y_1}{\partial x^2} = 8.$$

Since the second derivative is positive everywhere, it is positive at  $x = 3/8$ , and so by the second derivative test,  $y_1$  has a minimum at  $x = 3/8$ . Another approach to this question that does not require the second derivative test would be to recognize that this function is a convex quadratic, and so must have a minimum. Next, note that

$$y_2' = 12x^3 - 6x^2 = 0 \implies x = \frac{1}{2}$$

and

$$y_2'' = 36x^2 - 12x.$$

Evaluating the second derivative of  $y_2$  at  $x = 1/2$  gives:

$$y_2''(1/2) = 36 - 12 = 24 > 0,$$

and so we can conclude that  $y_2$  has a minimum at  $x = 1/2$ . Next, we have

$$y_3' = 4 - \frac{1}{2\sqrt{1-x}} = 0 \implies x = \frac{63}{64},$$

and

$$y_3'' = \frac{-1}{4(1-x)^{3/2}},$$

and so  $y_3''(\frac{63}{64}) < 0$ , so  $x$  is a maximum of  $y_3$ . Finally,

$$y_4' = 1 - x^{-2} = 0 \implies x = \pm 1$$

and

$$y_4'' = 2x^{-3}.$$

Therefore,  $y_4''(1) < 0$  and  $y_4''(-1) > 0$ , so  $y_4$  has a minimum at  $y = 1$  and a maximum at  $y = -1$ . In matplotlib, we can plot these functions to verify:

```

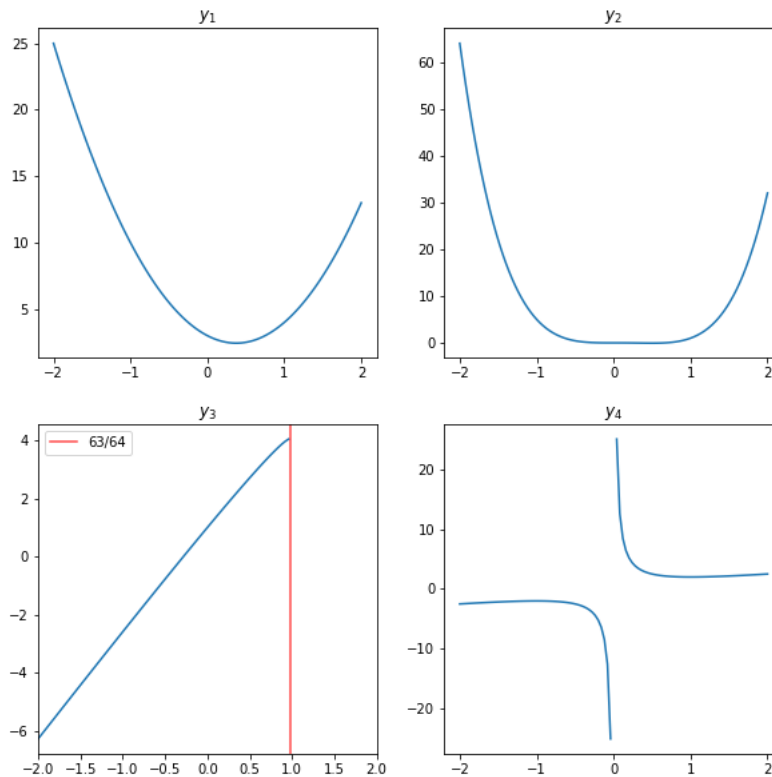
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 y1 = lambda x: 4*x**2 - 3*x + 3
5 y2 = lambda x: 3*x**4 - 2*x**3
6 y3 = lambda x: 4*x + np.sqrt(1-x)
7 y4 = lambda x: x + x**(-1)
8
9 x = np.linspace(start=-2, stop=2, num=101)
10
11 fig, axs = plt.subplots(2,2, figsize=(10,10))

```

```

12     axs[0,0].plot(x, y1(x))
13     axs[0,0].set_title("$y_1$")
14     axs[0,1].plot(x, y2(x))
15     axs[0,1].set_title("$y_2$")
16     axs[1,0].plot(x[x<1], y3(x[x<1]))
17     axs[1,0].axvline(63/64, color='red', alpha=0.7, label='63/64')
18     axs[1,0].legend()
19     axs[1,0].set_xlim(-2,2)
20     axs[1,0].set_title("$y_3$")
21     axs[1,1].plot(x, y4(x))
22     axs[1,1].set_title("$y_4$")
23     plt.savefig("4funs.png")
24     plt.show()
25

```



## Question 2. (Probability Review)

- (a) A manufacturing company has two retail outlets. It is known that 20% of potential customers buy

products from Outlet I alone, 10% buy from both I and II, and 40% buy from neither. Let  $A$  denote the event that a potential customer, randomly chosen, buys from outlet I, and  $B$  the event that the customer buys from outlet II. Compute the following probabilities:

$$P(A), \quad P(B), \quad P(A \cup B), \quad P(\bar{A}\bar{B})$$

**Solution:**

Hint: Draw a Venn diagram.

$$P(A) = 0.3$$

$$P(B) = 0.4$$

$$P(A \cup B) = P(A) + P(B) - P(AB) = 0.3 + 0.4 - 0.1 = 0.6$$

$$P(\bar{A}\bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 0.4.$$

Note that we have made regular use of the fact that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

and also made use of De Morgan's law which states that for any two sets  $A, B$ :

$$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \text{and} \quad \overline{A \cap B} = \bar{A} \cup \bar{B}.$$

Also note that it is customary to write:

$$A \cap B = AB.$$

- (b) Let  $X, Y$  be two discrete random variables, with joint probability mass function  $P(X = x, Y = y)$  displayed in the table below:

|     | $y$ |     |      |      |
|-----|-----|-----|------|------|
|     | 1   | 2   | 3    |      |
| $x$ | 1   | 1/6 | 1/12 | 1/12 |
|     | 2   | 1/6 | 0    | 1/6  |
|     | 3   | 0   | $r$  | 0    |

Compute the following quantities:

- (i)  $r$
- (ii)  $P(X = 2, Y = 3)$
- (iii)  $P(X = 3)$  and  $P(X = 3|Y = 2)$
- (iv)  $\mathbb{E}[X]$ ,  $\mathbb{E}[Y]$  and  $\mathbb{E}[XY]$
- (v)  $\mathbb{E}[X^2]$ ,  $\mathbb{E}[Y^2]$
- (vi)  $\text{Cov}(X, Y)$
- (vii)  $\text{Var}(X)$  and  $\text{Var}(Y)$
- (viii)  $\text{Corr}(X, Y)$
- (ix)  $\mathbb{E}[X + Y]$ ,  $\mathbb{E}[X + Y^2]$ ,  $\text{Var}(X + Y)$  and  $\text{Var}(X + Y^2)$ .

**Solution:**

- (i) Since this is a probability mass function, the probabilities must sum to 1, and so it follows that  $r = 1/3$ .
- (ii) We can just read this off the table  $P(X = 2, Y = 3) = 1/6$ .
- (iii) To compute  $P(X = 3)$ , we use the law of total probability, which allows us to sum over all possibilities for  $(X, Y)$  when  $X = 3$ , we can therefore write:

$$\begin{aligned}
 P(X = 3) &= \sum_{y=1}^3 P(X = 3, Y = y) \\
 &= P(X = 3, Y = 1) + P(X = 3, Y = 2) + P(X = 3, Y = 3) \\
 &= 1/3.
 \end{aligned}$$

Note that we are basically just adding the values in the row corresponding to  $X = 3$ . Then, to compute the conditional probability, we can use Bayes rule:

$$\begin{aligned}
 P(X = 3|Y = 2) &= \frac{P(X = 3, Y = 2)}{P(Y = 2)} \\
 &= \frac{P(X = 3, Y = 2)}{P(X = 1, Y = 2) + P(X = 2, Y = 2) + P(X = 3, Y = 2)} \\
 &= \frac{1/3}{1/12 + 0 + 1/3} \\
 &= \frac{1/3}{5/12} \\
 &= \frac{4}{5}.
 \end{aligned}$$

- (iv) Recall the formula for expectation is

$$\mathbb{E}[X] = \sum_x xP(X = x)$$

where the summation is over all possible values of  $X$ . So, for our case we have

$$\begin{aligned}
 \mathbb{E}[X] &= \sum_{x=1}^3 xP(X = x) \\
 &= 1 \times P(X = 1) + 2 \times P(X = 2) + 3 \times P(X = 3) \\
 &= 1 \times \frac{1}{3} + 2 \times \frac{1}{3} + 3 \times \frac{1}{3} \\
 &= 2.
 \end{aligned}$$

Similarly, we can show that

$$\mathbb{E}[Y] = \frac{23}{12}.$$

To compute  $XY$ , we can apply the same definition, which tells us that

$$\mathbb{E}[XY] = \sum_{x,y} xyP(X = x, Y = y),$$

so we are now summing over all possible value taken by  $X$  and  $Y$ :

$$\begin{aligned}\mathbb{E}[XY] &= 1 \times 1 \times P(X = 1, Y = 1) + 1 \times 2 \times P(X = 1, Y = 2) + 1 \times 3 \times P(X = 1, Y = 3) \\ &\quad + 2 \times 1 \times P(X = 2, Y = 1) + 2 \times 2 \times P(X = 2, Y = 2) + 2 \times 3 \times P(X = 2, Y = 3) \\ &\quad + 3 \times 1 \times P(X = 3, Y = 1) + 3 \times 2 \times P(X = 3, Y = 2) + 3 \times 3 \times P(X = 3, Y = 3) \\ &= \frac{47}{12}.\end{aligned}$$

(v) Similar to the previous question, we have

$$\mathbb{E}[X^2] = \sum_x x^2 P(X = x),$$

so this definition applies for any function of  $X$ . We therefore get

$$\begin{aligned}\mathbb{E}[X^2] &= 1^2 \times P(X = 1) + 2^2 P(X = 2) + 3^2 P(X = 3) \\ &= 1 \times \frac{1}{3} + 4 \times \frac{1}{3} + 9 \times \frac{1}{3} \\ &= \frac{14}{3}.\end{aligned}$$

(vi) Recall the formula for covariance:

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],$$

and using our results from previous questions we have

$$\text{Cov}(X, Y) = \frac{47}{12} - 2 \times \frac{23}{12} = \frac{1}{12}.$$

(vii) Recall that variance is defined by:

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

We will illustrate the usage of both definitions. First

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{14}{3} - (2)^2 = \frac{2}{3}.$$

Then, we compute the variance of  $Y$  using the standard definition of variance:

$$\begin{aligned}
 \text{Var}(Y) &= \mathbb{E}[(Y - \mathbb{E}[Y])^2] \\
 &= \mathbb{E}\left[\left(Y - \frac{23}{12}\right)^2\right] \\
 &= \sum_{y=1}^3 \left(y - \frac{23}{12}\right)^2 P(Y = y) \\
 &= \left(1 - \frac{23}{12}\right)^2 \times \frac{1}{3} + \left(2 - \frac{23}{12}\right)^2 \times \frac{5}{12} + \left(3 - \frac{23}{12}\right)^2 \times \frac{1}{4} \\
 &= 0.5763889.
 \end{aligned}$$

Note that you can use either approach, they will give identical results.

(viii) We denote the standard deviation of  $X$  by  $\text{SD}(X)$ , then

$$\begin{aligned}
 \text{Corr}(X, Y) &= \frac{\text{Cov}(X, Y)}{\text{SD}(X)\text{SD}(Y)} \\
 &= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}},
 \end{aligned}$$

then simply plug in the values we computed earlier.

(ix) Recall that expectation is linear, regardless of whether the random variables are dependent or independent, so we have

- (a)  $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- (b)  $\mathbb{E}[X + Y^2] = \mathbb{E}[X] + \mathbb{E}[Y^2]$

Note however that variance is **not** linear in general. We have

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

In general, for constants  $a, b$  and random variables  $X, Y$ ,

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).$$

### Question 3. (Linear Algebra Review)

(a) Write down the dimensions of the following objects:

$$A = \begin{bmatrix} 1 & 3 & 1 & 0 & 2 \\ 1 & 1 & 4 & 1 & 2 \\ 1 & 1 & 1 & 5 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 3 \\ 3 \\ 2 \end{bmatrix}, \quad A^T$$



**Solution:**

$A$  is a matrix with 3 rows and 5 columns, so we say that  $A$  is a  $3 \times 5$  matrix, or equivalently  $A \in \mathbb{R}^{3 \times 5}$ . Similarly,  $b$  is a matrix with a single column, also called a vector, and so  $b \in \mathbb{R}^{6 \times 1}$  or sometimes shortened to  $b \in \mathbb{R}^6$ .  $A^T$  is the transpose of  $A$ . To find the tranpose of  $A$ , we take element  $(i, j)$  in  $A$ , and make that element  $(j, i)$  in  $A^T$ . In this case we get

$$A^T = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 1 & 1 \\ 1 & 4 & 1 \\ 0 & 1 & 5 \\ 2 & 2 & 2 \end{bmatrix} \in \mathbb{R}^{5 \times 3},$$

note that the dimension of the transpose is just the reverse of the dimension of the original object. In Python, we can do:

```

1      import numpy as np
2      A = np.array([[1,3,1,0,2], [1,1,4,1,2], [1,1,1,5,2]])
3      dimA = A.shape           # dimension of A
4      AT = A.T                 # A transpose
5      dimAT = AT.shape         # dimension of A transpose
6

```

(b) Consider the following objects:

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 2 & 1 \\ 6 & 4 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 4 \\ 1 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 7 & 3 & 3 \\ 2 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}, \quad D = \begin{bmatrix} 4 & 2 \\ 4 & 6 \\ 1 & 3 \end{bmatrix}, \quad u = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad v = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$$

Compute the following:

- (i)  $AB$  and  $BA$
- (ii)  $AC$  and  $CA$
- (iii)  $AD$  and  $DA$
- (iv)  $DC$  and  $CD$  and  $D^T C$
- (v)  $Bu$  and  $uB$
- (vi)  $Au$
- (vii)  $Av$  and  $vA$
- (viii)  $Av + Bv$

**Solution:**

Before computing any of the above, we recall that for any two matrices  $X \in \mathbb{R}^{n \times p}$  and  $Z \in \mathbb{R}^{q \times m}$ , their matrix product  $XZ$  exists if and only if  $p = q$ , and the resulting matrix will have dimension  $n \times m$ .

- (i) Since  $A \in \mathbb{R}^{3 \times 3}$  and  $B \in \mathbb{R}^{2 \times 2}$  the matrix product  $AB$  is not defined, and neither is  $BA$ .

(ii) Both  $A$  and  $C$  are  $3 \times 3$  matrices, so their product exists, and is given by

$$AC = \begin{bmatrix} 21 & 14 & 14 \\ 20 & 10 & 10 \\ 56 & 28 & 28 \end{bmatrix}$$

Further,  $CA$  is also defined and given by:

$$CA = \begin{bmatrix} 31 & 39 & 40 \\ 10 & 12 & 12 \\ 18 & 18 & 16 \end{bmatrix}$$

So note that in general,  $AC \neq CA$ , so order matters. In fact, a matrix product might exist in one order, but not the other, as we will see below.

(iii)  $A \in \mathbb{R}^{3 \times 3}$  and  $D \in \mathbb{R}^{3 \times 2}$ , in this case,  $p = q = 3$ , and so the matrix product is defined, note that the resulting matrix will be of dimension  $3 \times 2$  and is given by:

$$AD = \begin{bmatrix} 20 & 32 \\ 17 & 19 \\ 43 & 45 \end{bmatrix}$$

However,  $DA$  is not defined, since now  $p = 2$  but  $q = 3$ .

(iv)  $DC$  is not defined, but  $CD$  is, and has dimension  $3 \times 2$ .  $D^T C$  is also defined and has dimension  $2 \times 3$ .

(v)  $Bu$  exists and has dimension  $2 \times 1$

$$Bu = \begin{bmatrix} 14 \\ 4 \end{bmatrix}$$

but  $uB$  of course does not exist.

(vi)  $Au$  does not exist.

(vii)  $Av$  exists, but  $vA$  does not.

(viii)

$$Av + Cv = (A + C)v = \begin{bmatrix} 8 & 6 & 7 \\ 4 & 3 & 2 \\ 8 & 6 & 5 \end{bmatrix} v = \begin{bmatrix} 47 \\ 22 \\ 45 \end{bmatrix}$$

In Python, we could compute the above as follows:

```
1 import numpy as np
2 A = np.array([[1,3,4], [2,2,1], [6,4,3]])
3 B = np.array([[2,4], [1,1]])
4 C = np.array([[7,3,3], [2,1,1], [2,2,2]])
5 D = np.array([[4,2], [4,6], [1,3]])
6 u = np.array([1,3]).reshape(2,1) # ensures correct shape for vector
```

```

7      v = np.array([2,4,1]).reshape(3,1)
8
9      AB = A @ B      # will not work
10     BA = B @ A      # will not work
11     AC = A @ C
12     CA = C @ A
13     AD = A @ D
14     DA = D @ A      # will not work
15     DC = D @ C      # will not work
16     CD = C @ D
17     DTC = D.T @ C
18     Bu = B @ u
19     A@v + C@v
20

```

(c) Consider the following objects:

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 2 & 1 \\ 6 & 4 & 3 \end{bmatrix}, \quad u = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad v = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}, \quad w = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}.$$

Compute the following:

- (i)  $\|u\|_1, \|u\|_2, \|u\|_2^2, \|u\|_\infty$
- (ii)  $\|v\|_1, \|v\|_2, \|v\|_2^2, \|v\|_\infty$
- (iii)  $\|v + w\|_1, \|v + w\|_2, \|v + w\|_\infty$
- (iv)  $\|Av\|_2, \|A(v - w)\|_\infty$

**Solution:**

Recall that for a general vector  $x \in \mathbb{R}^n$ , the  $p$ -norm is defined by

$$\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}$$

and for  $p = \infty$ , this becomes:

$$\|x\|_\infty = \max_{i=1, \dots, n} |x_i|,$$

i.e. the  $\infty$ -norm is just the largest (absolute) element in the vector. Note that in the special case  $p = 2$ , we have  $\|x\|_2^2 = x^T x$ . Let's work through (ii) explicitly:

$$\|v\|_1 = \sum_{i=1}^3 |v_i| = |2| + |4| + |1| = 7$$

$$\|v\|_2 = \sqrt{\sum_{i=1}^3 |v_i|^2} = \sqrt{2^2 + 4^2 + 1^2} = \sqrt{21}$$

$$\|v\|_2^2 = v^T v = 21$$

$$\|v\|_\infty = \max\{|2|, |4|, |1|\} = 4$$

Note that a norm is a way of measuring the size of an object, and different choices of  $p$  correspond to different ways of measuring the size. In Python we would compute:

```
1 v_1norm = np.linalg.norm(v, ord=1)
2 v_2norm = np.linalg.norm(v, ord=2)
3 v_2norm_sq = np.linalg.norm(v, ord=2)**2
4 v_infnorm = np.linalg.norm(v, ord=np.inf)
5
```

(d) Consider the following vectors in  $\mathbb{R}^2$

$$u = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad w = \begin{bmatrix} -1 \\ 1/2 \end{bmatrix}$$

Compute the dot products between all pairs of vectors. Note that the dot product may be written using the following equivalent forms:

$$\langle x, y \rangle = x \cdot y = x^T y.$$

Then compute the angle between the vectors and plot.

**Solution:**

$$\langle u, v \rangle = 1 \times 1 + 2 \times 1 = 3$$

$$\langle u, w \rangle = 0$$

$$\langle v, w \rangle = -1/2$$

To compute the angle between vectors  $x$  and  $y$ , which we will label as  $\theta_{xy}$ , we recall the following important formula (we will explore this result in more depth in the next part):

$$\langle x, y \rangle = \|x\|_2 \|y\|_2 \cos(\theta_{xy}),$$

and so rearranging yields:

$$\theta_{xy} = \arccos\left(\frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2}\right).$$

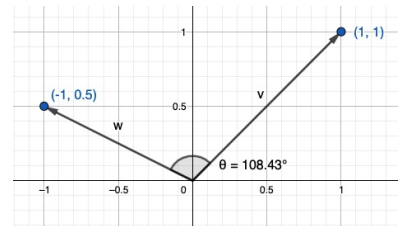
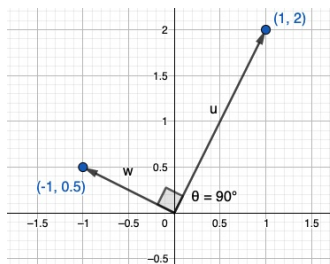
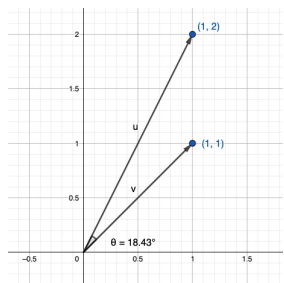
Then, for example, we have

$$\theta_{uv} = \arccos\left(\frac{3}{\sqrt{5} \times \sqrt{2}}\right) = 18.43^\circ.$$

We similarly have

$$\theta_{uw} = 90^\circ, \quad \theta_{vw} = 108.43^\circ.$$

The following plots (created using [Geogebra](#)) depict the three scenarios:



In python, we can do:

```

1  u = np.array([1,2])
2  v = np.array([1,-1])
3  w = np.array([-1,-1])
4  dotuv = np.dot(u, v)
5
6  def angle_bw(x,y):
7      dotprod = np.dot(x,y)
8      norm_x = np.linalg.norm(x, ord=2)
9      norm_y = np.linalg.norm(y, ord=2)
10     # arccos returns radians, we want degrees
11     return np.degrees(np.arccos(dotprod/(norm_x*norm_y)))
12
13  theta_uv = angle_bw(u,v)
14

```

- (e) Dot products are extremely important in machine learning, explain what it means for a dot product to be zero, positive or negative.

### Solution:

Recall the following angle types:

1. acute angle:  $\theta \in [0^\circ, 90^\circ)$
2. right angle:  $\theta = 90^\circ$
3. obtuse angle:  $\theta \in [90^\circ, 180^\circ)$

Continuing from the previous question, we saw that when a dot product was positive, the resulting angle was acute, when the dot-product was zero, then the angle was a right angle, and when the dot-product was negative, the result was an obtuse angle. In machine learning, we use dot-products to measure the strength of relationship between two objects (where the objects are represented as vectors). If we take the dot product between two vectors and it is positive, then we know that the objects point in the same direction (acute). The right angle is the special case where the objects neither agree or disagree, and the obtuse angle tells us they are pointing in different directions. We will often just be concerned with the **sign** of the dot product (is it positive or negative?) and not the exact angle.

(f) Consider the  $2 \times 2$  matrix:

$$A = \begin{bmatrix} 1 & 3 \\ 4 & 1 \end{bmatrix}$$

Compute the inverse of  $A$ .

**Solution:**

Recall that the inverse matrix of  $A$  is denoted  $A^{-1}$ , and is defined as the matrix that satisfies  $AA^{-1} = I$ , where  $I$  is the identity matrix:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

. Now, for the  $2 \times 2$  case, there is a simple formula for the determinant. For a general  $2 \times 2$  matrix

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then the inverse is

$$X^{-1} = \frac{1}{\det(X)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

where  $\det$  denotes the matrix determinant. Using this formula, we have

$$A^{-1} = \frac{1}{1 \times 1 - 4 \times 3} \begin{bmatrix} 1 & -3 \\ -4 & 1 \end{bmatrix} = -\frac{1}{11} \begin{bmatrix} 1 & -3 \\ -4 & 1 \end{bmatrix} = \begin{bmatrix} -1/11 & 3/11 \\ 4/11 & 1/11 \end{bmatrix}$$

To check that this is indeed the inverse, we can verify that  $A^{-1}A = I$ . Note that in Python, we can do the following:

```
1 import numpy as np
2 from numpy.linalg import det, inv
3
4 A = np.array([[1,3], [4,1]])
5 detA = det(A)
6 invA = inv(A)
7
```

(g) Consider the  $2 \times 2$  matrix

$$A = \begin{bmatrix} 3 & 3 \\ 4 & 4 \end{bmatrix}$$

Compute its inverse  $A^{-1}$ .

**Solution:**

Note that in this case,  $\det(A) = 12 - 12 = 0$ , so in order to use the formula from the previous part, we would have to divide by zero. In other words, the inverse does not exist in this case.

(h) Let  $X$  be a matrix (of any dimension), show that  $X^T X$  is always symmetric.

**Solution:**

Recall that a symmetric matrix is a matrix that is equal to its transpose, i.e.  $A$  is symmetric if and only if  $A = A^T$ . Now, it follows easily that

$$(X^T X)^T = X^T (X^T)^T = X^T X,$$

so  $X^T X$  is symmetric, and this holds for any matrix  $X$ . Note that we used two results in the previous calculation, namely that the transpose of a transpose just gives back the original matrix:

$$(A^T)^T = A$$

and that taking a transpose of a product reverses the order:

$$(AB)^T = B^T A^T.$$