Refreshing the Rows and Columns: Linear Algebra Review

1.1 Basic Operations

Use the definitions below

α=2 $\mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$, $\mathbf{z} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{A} = \begin{bmatrix} 3 & 2 & 2 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$

and use x_i to denote element i of vector x. Evaluate the following expressions:

1. $\sum_{i=1}^{n} x_i y_i$ (inner product): **The answer is 14**
2. $\sum_{i=1}^{n} x_i z_i$ (inner product between orthogonal vectors): **The answer is 0. The dot product of 2 orthogonal vectors is 0, as can be shown in the calculation or sum product.**
3. α(x + y) (vector addition and scalar multiplication): **The answer is $\begin{bmatrix} 6 \\ 10 \\ 14 \end{bmatrix}$, i.e. the scalar times every element in the vector that results from the dot product.**
4. $||\mathbf{x}||$ (Euclidean norm of x): **The answer is $\sqrt{5}$, i.e. the squareroot of the sum of the squares of all the elements in the vector.**
5. $\mathbf{x}^T$ (vector transpose): **The answer is $\begin{bmatrix} 0 & 1 & 2 \end{bmatrix}$**
6. Ax (matrix-vector multiplication): **The answer is $\begin{bmatrix} 6 \\ 5 \\ 7 \end{bmatrix}$**
7. $\mathbf{x}^T$ Ax (quadratic form): **The answer is $\begin{bmatrix} 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 5 \\ 7 \end{bmatrix} = 19$**

Note, you do not need to show your work

1.2 Matrix Algebra Rules

Assume that {x, y, z} are n × 1 column vectors and { A, B, C } are n × n real-valued matrices, and I is the identity matrix of appropriate size. State whether each of the below is true in general (you do not need to show your work).

1. $\mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{n} x_i y_i$: **True. This is the definition of dot product.**
2. $\mathbf{x}^\top \mathbf{x} = ||\mathbf{x}||^2$: **True. The dot product of a vector and itself is equalt to its L2 or Euclidean norm.**
3. $\mathbf{x}^\top \mathbf{x} = \mathbf{x}\mathbf{x}^\top$: **False. The left side of the equation results in a scalar (1x1 matrix) while the right hand side results in an nxn matrix.**
4. $(x-y)^\top (y-x) = ||x||^2 - 2x^\top y + ||y||^2$: **False. Expanding the left side of the equation simplifies to $2x^\top y - ||x||^2 - ||y||^2$. This is different from the right side of the equation.**
5. $AB = BA$: **False. Matrix multiplication is generally not commutative.**

6. $A(B+C) = AB + AC$: **True. matrix multiplication is distributive over addition.**
7. $(AB)^\top = A^\top B^\top$: **False, $(AB)^\top$ will be equal to $B^\top A^\top$. The transpose of a product of matrices is the same as the product of the transposes in reverse order.**
8. $\mathbf{x}^\top \mathbf{A}\mathbf{y} = \mathbf{y}^\top \mathbf{A}^\top \mathbf{x}$: **True. $\mathbf{x}^\top \mathbf{A}\mathbf{y} = (\mathbf{x}^\top \mathbf{A}\mathbf{y})^\top$ = $\mathbf{y}^\top \mathbf{A}^\top (\mathbf{x}^\top)^\top = \mathbf{y}^\top \mathbf{A}^\top \mathbf{x}$**
9. $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$ if the columns of $\mathbf{A}$ are orthonormal: **True. If A's columns are orthonormal, that means A is an orthogonal matrix. Therefore, $A^\top = A^{-1}$, and $A^\top A = I$.**

1.3 Matrix Operations

Let $\mathbf{B} = \begin{bmatrix} -1 & 0 & -1 \\ 0 & 2 & -1 \\ -1 & 1 & 3 \end{bmatrix}$

1. Is $\mathbf{B}$ invertible? If so, find $\mathbf{B}^{-1}$. **Yes, B is invertible. This can be proven by calculating the determinant of B. If the determinant is zero, the matrix is said to be singualr, and non-invertible. If the determinant if NOT zero, the the matrix is invertible. See the code below that shows that B is invertible.**
2. Is $\mathbf{B}$ diagonalizable? If so, find its diagonalization. **Yes, $\mathbf{B}$ is diagonalizable. The matrix $\mathbf{B}$ is diagonalizable if it is similar to a diagonal matrix $\mathbf{D}$, such that there exists an invertible matrix $\mathbf{P}$ where $\mathbf{B} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. The diagonal entries of $\mathbf{D}$ are the eigenvalues of $\mathbf{B}$, and the columns of $\mathbf{P}$ are the corresponding eigenvectors. A key condition for diagonalizability is that the matrix must have a complete set of linearly independent eigenvectors. For an $n \times n$ matrix, it must have n linearly independent eigenvectors. The code below shows that $\mathbf{B}$ has n linearly independent eigenvectors and is therefore diagonalizable. the resulting diagonal matrix $\mathbf{D}$ is also shown below.**

In [12]:
```
# Show whether B is invertible or not, based on the value of the determin
import numpy as np

B = np.array([[-1, 0, -1],
              [0, 2, -1],
              [-1, 1, 3]])
det_B = np.linalg.det(B)
print(f'The determinant of B is: {det_B: 0.4f}')
print('Is B invertible (True/False)? ', det_B != 0)
```

```
The determinant of B is: -9.0000
Is B invertible (True/False)?  True
```

In [19]:
```
# Showing eigenvalues of B
eigenvalues,  eigenvectors = np.linalg.eig(B)

print("Eigenvalues of B:")
print(eigenvalues)

# Since second and third eigenvalues are the same, check for linear indep
eigenvector_2 = eigenvectors[:, 1]
eigenvector_3 = eigenvectors[:, 2]
```

```python
# Create matrix of eigenvectors 2 and 3 and check their rank
eigenvectors_subset = np.column_stack((eigenvector_2, eigenvector_3))
rank = np.linalg.matrix_rank(eigenvectors_subset)
print('Eigenvectors 2 and 3 are linearly independent (True/False)? ', ran

# Diagonalize B
D = np.diag(eigenvalues)
print('Diagonal matrix D:')
print(D)


# Verify diagonalization
P = eigenvectors
P_inv = np.linalg.inv(P)
B_reconstructed = P @ D @ P_inv
print('Is B_reconstructed equal to B (True/False)? ', np.allclose(B, B_re
```

```
Eigenvalues of B:
[-1.22069282+0.j          2.61034641+0.74763217j  2.61034641-0.74763217j]
Eigenvectors 2 and 3 are linearly independent (True/False)?  True
Diagonal matrix D:
[[-1.22069282+0.j          0.        +0.j          0.        +0.j        ]
 [ 0.        +0.j          2.61034641+0.74763217j  0.        +0.j        ]
 [ 0.        +0.j          0.        +0.j          2.61034641-0.74763217
j]]
Is B_reconstructed equal to B (True/False)?  True
```

2 Taking Chances: probability Review

2.1 Basic probability

1. You are offered the opportunity to play the following game: your opponent rolls 2 regular 6-sided dice. If the difference between the two rolls is at least 3, you win 15 dollars. Otherwise, you get nothing. What is a fair price for a ticket to play this game once? In other words, what is the expected value of playing the game?
   Answer: There a 36 possible outcomes from folling the 2 dices. Of the 36, 12 are winning outcomes, and the rest are losing outcomes. The fair value of this game can be calculated as the probability of winning times the reward for winning, plus the probability of losing times the reward for losing.
   $E = (\text{Value of Winning} \times P(\text{win})) + (\text{Value of Losing} \times P(\text{lose}))$
   $E = (15 \times \frac{1}{3}) + (0 \times \frac{2}{3})$ $E = 5 + 0 = 5$

2. Consider two events A and B such that $P(A \cap B) = 0$ (they are mutually exclusive). If $P(A) = 0.4$ and $P(A \cup B) = 0.95$, what is $P(B)$? **Note:** $P(A \cap B)$ means "probability of A and B" while $P(A \cup B)$ means "probability of A or B". It may be helpful to draw a Venn diagram. **Answer: For any 2 events,** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$**. For mutually exclusive events,** $P(A \cap B) = 0$**, therefore** $P(A \cup B) = P(A) + P(B)$**. With the values substituded, this becomes** $0.95 = 0.4 + P(B)$**. Hence,** $P(B) = 0.55$**.**

3. Instead of assuming that A and B are mutually exclusive ($P(A \cap B) = 0$), what is the answer to the previous question if we assume that A and B are independent?
   Answer: For independent events, the probability of both events occuring

$P(A \cap B)$ **is** $P(A) \times P(B)$**. The general rule** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ **becomes** $0.95 = 0.4 + P(B) - (0.4 \times P(B)) = 0.4 + P(B) - 0.4P(B) = 0.4 + 0.6P(B)$ **. When you solve** $0.95 = 0.4 + 0.6P(B)$ **for** $P(B)$**, you end up with** $P(B) = \frac{0.55}{0.6}$

2.2 Expectations and Variance

Suppose we have two coins. Coin C1 comes up heads with probability 0.3 and coin C2 comes up heads with probability 0.9. We repeat this process 3 times:

- Choose a coin with equal probability
- Flip the coin once

Suppose X is the number of heads after 3 flips.

1. What is $E[X]$?

The probabilities for each number of heads are:

- **For X=0:** $P(X = 0|C1) = \binom{3}{0}(0.3)^0(0.7)^3 = 1 \times 1 \times 0.343 = 0.343$
  $P(X = 0|C2) = \binom{3}{0}(0.9)^0(0.1)^3 = 1 \times 1 \times 0.001 = 0.001$
  $P(X = 0) = 0.343 \times 0.5 + 0.001 \times 0.5 = 0.1715 + 0.0005 = 0.172$
- **For X=1:** $P(X = 1|C1) = \binom{3}{1}(0.3)^1(0.7)^2 = 3 \times 0.3 \times 0.49 = 0.441$
  $P(X = 1|C2) = \binom{3}{1}(0.9)^1(0.1)^2 = 3 \times 0.9 \times 0.01 = 0.027$
  $P(X = 1) = 0.441 \times 0.5 + 0.027 \times 0.5 = 0.2205 + 0.0135 = 0.234$
- **For X=2:** $P(X = 2|C1) = \binom{3}{2}(0.3)^2(0.7)^1 = 3 \times 0.09 \times 0.7 = 0.189$
  $P(X = 2|C2) = \binom{3}{2}(0.9)^2(0.1)^1 = 3 \times 0.81 \times 0.1 = 0.243$
  $P(X = 2) = 0.189 \times 0.5 + 0.243 \times 0.5 = 0.0945 + 0.1215 = 0.216$
- **For X=3:** $P(X = 3|C1) = \binom{3}{3}(0.3)^3(0.7)^0 = 1 \times 0.027 \times 1 = 0.027$
  $P(X = 3|C2) = \binom{3}{3}(0.9)^3(0.1)^0 = 1 \times 0.729 \times 1 = 0.729$
  $P(X = 3) = 0.027 \times 0.5 + 0.729 \times 0.5 = 0.0135 + 0.3645 = 0.378$

Therefore,

$$E[X] = \sum x \cdot P(X = x)$$

$$E[X] = (0 \times 0.172) + (1 \times 0.234) + (2 \times 0.216) + (3 \times 0.378)$$

$$E[X] = 0 + 0.234 + 0.432 + 1.134$$

$$E[X] = 1.8$$

2. What is $Var[X]$?

$$Var[X] = E[X^2] - (E[X])^2$$

Calculating $E[X^2]$, the expected value of X squared:

$E[X^2] = \sum x^2 \cdot P(X = x)$

$E[X^2] = (0^2 \times 0.172) + (1^2 \times 0.234) + (2^2 \times 0.216) + (3^2 \times 0.378)$

$E[X^2] = (0 \times 0.172) + (1 \times 0.234) + (4 \times 0.216) + (9 \times 0.378)$

$E[X^2] = 0 + 0.234 + 0.864 + 3.402$

$E[X^2] = 4.5$

Therefore, $Var[X] = E[X^2] - (E[X])^2$ is:

$Var[X] = 4.5 - (1.8)^2$

$Var[X] = 4.5 - 3.24$

$Var[X] = 1.26$

3. Based on the number of heads we get, we earn $Y = \frac{1}{2+X}$ dollars. What is $E[Y]$?

$E[Y] = \sum y \cdot P(Y = y)$

The possible values for Y are:

- If X=0, $Y = \frac{1}{2+0} = 0.5$, with $P(Y = 0.5) = P(X = 0) = 0.172$
- If X=1, $Y = \frac{1}{2+1} = \frac{1}{3}$, with $P(Y = 1/3) = P(X = 1) = 0.234$
- If X=2, $Y = \frac{1}{2+2} = 0.25$, with $P(Y = 0.25) = P(X = 2) = 0.216$
- If X=3, $Y = \frac{1}{2+3} = 0.2$, with $P(Y = 0.2) = P(X = 3) = 0.378$

Therefore, $E[Y]$:

$E[Y] = (0.5 \times 0.172) + (\frac{1}{3} \times 0.234) + (0.25 \times 0.216) + (0.2 \times 0.378)$

$E[Y] = 0.086 + 0.078 + 0.054 + 0.0756$

$E[Y] = 0.2936$

2.3 A Variance Paradox

For independent identically distributed (i.i.d.) random variables $X_1, \ldots, X_n$, each with distribution F and variance σ^2. We know that $\mathrm{Var}[X_1 + \cdots + X_n] = n\sigma^2$. On the other hand, if $X \sim F$, then $\mathrm{Var}[X + X] = \mathrm{Var}[2X] = 4\sigma^2$. Is there a contradiction here? Explain.

There is no contradiction, as the two scenarios are different:

- In the first case ($\mathrm{Var}[X_1 + \cdots + X_n]$), the variables being added (X_1, X_2, etc.) are independent. The variance of a sum is only the sum of variances when the variables are independent.
- In the second case ($\mathrm{Var}[X + X]$ or $\mathrm{Var}[2X]$), the terms being added are the same random variable X. They are not independent. The rule that applies here is the rule for the variance of a scalar multiple.

calculus Review

3.1 One-variable derivatives

Answer the following questions. You do not need to show your work.

1. Find the derivative of the function $f(x) = 3x^2 - 2x + 5$.: **The answer is** $6x - 2$
2. Find the derivative of the function $f(x) = x(1 - x)$.: **The answer is** $1 - 2x$
3. Let $p(x) = \frac{1}{1+\exp(-x)}$ for $x \in \mathbb{R}$. Compute the derivative of the function $f(x) = x - \log(p(x))$ and simplify it by using the function $p(x)$.: **The answer is** $p(x)$

Note: In this course, we will use $\log(x)$ to mean the "natural" logarithm of x, so that $\log(\exp(1)) = 1$. Also, observe that $p(x) = 1 - p(-x)$ for the final part.

3.2 Multi-variable derivative

Compute the gradient $\nabla f(\mathbf{x})$ of each of the following functions. You do not need to show your work.

1. $f(\mathbf{x}) = x_1^2 + \exp(x_2)$ where $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \in \mathbb{R}^2$.: **Answer:** $\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 \\ \exp(x_2) \end{bmatrix}$
2. $f(\mathbf{x}) = \exp(x_1 + x_2 x_3)$ where $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \in \mathbb{R}^3$.: **Answer:** $$\nabla f(\mathbf{x}) = \exp(x_1 + x_2 x_3) \begin{bmatrix} 1 \\ x_3 \\ x_2 \end{bmatrix}$$
3. $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$ where $\mathbf{x} \in \mathbb{R}^2$ and $\mathbf{a} \in \mathbb{R}^2$.: **Answer:** $\nabla f(\mathbf{x}) = \mathbf{a}$
4. $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x}$ where $\mathbf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ and $\mathbf{x} \in \mathbb{R}^2$.: **Answer: Answer:** $$\nabla f(\mathbf{x}) = \begin{bmatrix} 4x_1 - 2x_2 \\ -2x_1 + 2x_2 \end{bmatrix} \text{ or } 2\mathbf{A}\mathbf{x}$$

5. $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$ where $\mathbf{x} \in \mathbb{R}^d$.: **Answer:** $\nabla f(\mathbf{x}) = \mathbf{x}$

Hint: it is helpful to write out the linear algebra expressions in terms of summations.

4 Algorithms and data Structure Review

For these questions you may find it helpful to review these notes or this Wiki page on big-O notation. Now, answer the following questions using big-O notation You do not need to show your work.

1. What is the cost of running the merge-sort algorithm to sort a list of n numbers? **Answer:** $O(n \log n)$
2. What is the cost of finding the third-largest element of an unsorted list of n numbers? **Answer:** $O(n)$**. The kth largest element can be found in linear time with Quickselect algorithm.**
3. What is the cost of finding the smallest element greater than 0 in a *sorted* list with n numbers? **Answer:** $O(\log n)$**. Since the list is sorted, this can be done with binary search.**
4. What is the cost of finding the value associated with a key in a hash table with n numbers? (Assume the values and keys are both scalars.) **Answer:** $O(1)$
5. What is the cost of computing the matrix-vector product $\mathbf{Ax}$ when $\mathbf{A}$ is $n \times d$ and $\mathbf{x}$ is $d \times 1$? **Answer:** $O(nd)$**. Computing each element of the resulting** $n \times 1$ **vector involves** d **multiplications and** $d-1$ **additions. There are** n **elements, which leads to** nd **multiplications and additions.**
6. What is the cost of computing the quadratic form $\mathbf{x}^\top \mathbf{Ax}$ when $\mathbf{A}$ is $d \times d$ and $\mathbf{x}$ is $d \times 1$? **Answer:** $O(d^2)$**, quadratic complexity**
7. What is the cost of computing matrix multiplication $\mathbf{AB}$ when $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times d$? **Answer:** $O(mnd)$**. The resulting matrix is** $m \times d$**. Each element of the resulting matrix is the dot product of a row from A (size** n**) and a column from B (size** n**), which costs** $O(n)$**. There are** $m \times d$ **such elements, leading to a total cost of** $O(mnd)$

5 Programming

```
In [ ]:
```