Exercise 1

We can define the *edit distance* $d_{\text{edit}}(w, w') : \Sigma^2 \to \mathbb{R}$ as follows. (Let $w = w_1 \dots w_n$ and $w' = w'_1 \dots w'_m$)

$$d_{\text{edit}}(w, w') \mapsto \begin{cases} |w| & \text{if } |w'| = 0 \\ |w'| & \text{if } |w| = 0 \\ d_{\text{edit}}(w_2 \dots w_n, w'_2 \dots w'_m) & \text{if } w_1 = w'_1 \\ 1 + \min \begin{cases} d_{\text{edit}}(w_2 \dots w_n, w') \\ d_{\text{edit}}(w, w'_2 \dots w'_m) \\ d_{\text{edit}}(w_2 \dots w_n, w'_2 \dots w'_m) \end{cases} & \text{otherwise} \end{cases}$$

As this definition of d_{edit} works by removing at most the first character of each word, we can proof by induction the length of $x, y, z \in \Sigma$, that d_{edit} is a metric on Σ:

- Let $|x| = |y| = |z| = 0$. Therefore, also $x = y = z$.
  Then $0 \leq d_{\text{edit}} = 0$. Thus, Nonnegativity is given.
  Since $x = y$, also $d_{\text{edit}}(x, y) = d_{\text{edit}}(y, x)$. Thus, Symmetry is given.
  Since $x = y = z$, the Triangle Inequality $d_{\text{edit}}(x, z) \leq d_{\text{edit}}(x, y) + (d_{\text{edit}})(y, z) \Leftrightarrow 0 \leq 0 + 0$ is given.

- Let $x = x_1 \dots x_n$, $y = y_1 \dots y_m$, and $z = z_1 \dots z_o$, $n, m, o \geq 1$. For $x' = x_2 \dots x_n$, $y' = y_2 \dots y_m$, and $z' = z_2 \dots z_o$ Nonnegativity, Symmetry, and the Triangle Inequality of d_{edit} is given.

- Since $n, m \geq 1$, the second rule of Nonnegativity, namely $d_{\text{edit}}(x, y) \Leftrightarrow x = y$ does not apply here.
  Since all $d_{\text{edit}}(x', y'), d_{\text{edit}}(x', y), d_{\text{edit}}(x, y')$ are non-negative, by definition of d_{edit}, $d_{\text{edit}}(x, y)$ must be non-negative as well. Therefore, the Nonnegativity of d_{edit} is proven.

- If $x_1 = y_1$, then $d_{\text{edit}}(x, y) = d_{\text{edit}}(x', y') = d_{\text{edit}}(y', x') = d_{\text{edit}}(y, x)$
  If $x_1 \neq y_1$. Since there is a shortest path to convert x into y, we can denote the order of operations as $(op_1, \dots, op_p)$. The shortest path to convert y into x is also of the same length and can be denoted as $(anti(op_p), \dots, anti(op_1))$, where *anti* denotes the opposite to an operation and is defined as follows:

$$\begin{aligned} anti(ins_i^a(w_1 \dots w_{i-1} w_i w_{i+1} w_n)) &\coloneqq del_i(w_1 \dots w_{i-1} a w_{i+1} w_n) \\ anti(del_i(w_1 \dots w_{i-1} w_i w_{i+1} w_n)) &\coloneqq ins_i^{w_i}(w_1 \dots w_{i-1} w_{i+1} w_n) \\ anti(repl_i^a(w_1 \dots w_{i-1} w_i w_{i+1} w_n)) &\coloneqq repl_i^{w_i}(w_1 \dots w_{i-1} a w_{i+1} w_n) \end{aligned}$$

  Thus, Symmetry is proven.

- We can denote the path to convert x into y and the path to convert y into z by the order of operations $O_1 \coloneqq (op_1, \dots, op_p), O_2 \coloneqq (op'_1, \dots, op'_p)$. Since $p, p' \geq 0$, since d_{edit} is nonnegative, the shortest path of converting x to z must not be larger than $p + p'$, since $(op_1, \dots, op_p, op'_1, \dots, op'_p)$ is also a valid order of operations for that purpose. Thus, the Triangle Inequality is proven.

Since d_{edit} is nonnegative, symmetric, and fulfills the triangle inequality, d_{edit} is a metric.

Fabian Grob
Simon Michau
Til Mohr

Exercise 2

Result (see Appendix for code):

Classification: k=2 Manhattan Distance
Test (1, -2, 0): Prediction 1
Test (4, -0.5, 2): Prediction -1
Test (1, 1.5, -2.5): Prediction 0
Test (-2, -1, -2): Prediction 0
Test (-4, -1, -1): Prediction 0

Classification: k=3 Manhattan Distance
Test (1, -2, 0): Prediction 1
Test (4, -0.5, 2): Prediction -1
Test (1, 1.5, -2.5): Prediction 1
Test (-2, -1, -2): Prediction -1
Test (-4, -1, -1): Prediction 1

Classification: k=2 Euclidean Distance
Test (1, -2, 0): Prediction 1
Test (4, -0.5, 2): Prediction -1
Test (1, 1.5, -2.5): Prediction 1
Test (-2, -1, -2): Prediction 0
Test (-4, -1, -1): Prediction 0

Classification: k=3 Euclidean Distance
Test (1, -2, 0): Prediction 1
Test (4, -0.5, 2): Prediction -1
Test (1, 1.5, -2.5): Prediction 1
Test (-2, -1, -2): Prediction 1
Test (-4, -1, -1): Prediction -1

Further Reasoning

(a),(b)

Coordinate	Label	(1,-2,0)	(4,-0.5,2)	(1,1.5,-2.5)	(-2,-1,-2)	(-4,-1,-1)
(-4,-2.1,-1)	-1	6.1	12.6	10.1	4.1	1.1
(-3.6,-1.4,0.2)	1	5.3999	10.3	10.2	4.2	1.9999
(1,-0.2,-0.3)	1	2.1	5.6	3.9	5.5	6.5
(0.3,-0.5,-0.5)	1	2.7	6.2	4.7	4.3	5.3
(-2, -3.5, -1)	-1	5.5	12.0	9.5	3.5	4.5
(-4.2, -4, 0.2)	1	7.4	13.5	13.3999	7.4	4.4
(-1.3, -0.1, -3)	1	7.1999	10.7	4.4	2.6	5.6
(-0.7, 0.9, -0.7)	1	5.3	8.8	4.1	4.5	5.4999
(1, 2, 1.4)	1	5.4	6.1	4.4	9.4	10.4
(2.6, -1.5, 0.2)	1	2.3	4.2	7.3	7.3	8.2999
(2, 4.3, -0.7)	-1	8.0	9.5	5.6	10.6	11.6
(0.6, 0.4, 0.2)	-1	3.0	6.1	4.2	6.2	7.2
(2.9, -1.7, 3.6)	-1	5.8	3.9	11.2	11.2	12.2
(3.6, 0.4, -2.5)	-1	7.5	5.8	3.7	7.5	10.5
(1.2, 4, 1.2)	-1	7.4	8.1	6.4	11.3999	12.3999
(-1, 0.5, 0.5)	-1	5.0	7.5	6.0	5.0	6.0
(3, 2.7, 2.3)	-1	9.0	4.5	8.0	13.0	14.0
(4, -3, 2.2)	-1	6.2	2.7	12.2	12.2	13.2
(0.1, 0.1, 3.5)	-1	6.5	6.0	8.3	8.7	9.7
(2.8, 1.2, 2.4)	-1	7.4	3.3	7.0	11.4	12.4
Classification (k=2)		1	-1	0	0	0
Classification (k=3)		1	-1	1	-1	1

Table 1: Manhattan distance table for the 5 query points. Classifications for k=2 (a) and k=3 (b) are stated below.

(c),(d)

Coordinate	Label	(1,-2,0)	(4,-0.5,2)	(1,1.5,-2.5)	(-2,-1,-2)	(-4,-1,-1)
(-4, -2.1, -1)	-1	5.1	8.692	6.341	2.491	1.1
(-3.6, -1.4, 0.2)	1	4.643	7.862	6.071	2.749	1.326
(1, -0.2, -0.3)	1	1.825	3.792	2.780	3.539	5.111
(0.3, -0.5, -0.5)	1	1.729	4.465	2.913	2.791	4.357
(-2, -3.5, -1)	-1	3.5	7.348	6.020	2.692	3.201
(-4.2, -4, 0.2)	1	5.575	9.095	8.036	4.322	3.237
(-1.3, -0.1, -3)	1	4.231	7.297	2.846	1.516	3.478
(-0.7, 0.9, -0.7)	1	3.434	5.598	2.547	2.643	3.819
(1, 2, 1.4)	1	4.238	3.950	3.931	5.436	6.305
(2.6, -1.5, 0.2)	1	1.688	2.489	4.341	5.123	6.726
(2, 4.3, -0.7)	-1	6.417	5.859	3.475	6.766	8.011
(0.6, 0.4, 0.2)	-1	2.441	3.950	2.942	3.682	4.955
(2.9, -1.7, 3.6)	-1	4.082	2.282	7.145	7.473	8.322
(3.6, 0.4, -2.5)	-1	4.332	4.606	2.823	5.793	7.872
(1.2, 4, 1.2)	-1	6.122	5.360	4.469	6.743	7.541
(-1, 0.5, 0.5)	-1	3.240	5.315	3.741	3.082	3.674
(3, 2.7, 2.3)	-1	5.602	3.366	5.336	7.561	8.577
(4, -3, 2.2)	-1	3.852	2.507	7.165	7.592	8.845
(0.1, 0.1, 3.5)	-1	4.180	4.221	6.226	5.989	6.186
(2.8, 1.2, 2.4)	-1	4.386	2.118	5.228	6.873	7.914
Classification (k=2)		1	-1	1	0	0
Classification (k=3)		1	-1	1	1	-1

Table 2: Euclidian distance table for the 5 query points. Classifications for k=2 (c) and k=3 (d) are stated below.

Exercise 3

(a) $X_1 \land X_2 \land X_3$ is obviously in 1-CNF, thus also in 2-CNF. There exists no satisfiability equivalent formula in 2-DNF. For this, there would have to be at least two disjunctions of conjunctions of at most 2 literals. Thus, such a formula would be true even if one literal would be false.

(b) $X_1 \lor X_2 \lor X_3$ is obviously in 1-DNF, thus also in 2-DNF. There exists no satisfiability equivalent formula in 2-CNF. Since such a formula can only have 2 literals in each disjunction, there is no possibility of validating, that the third literal might be true. Therefore, such a formula is not true, when just one literal is true, the others false.

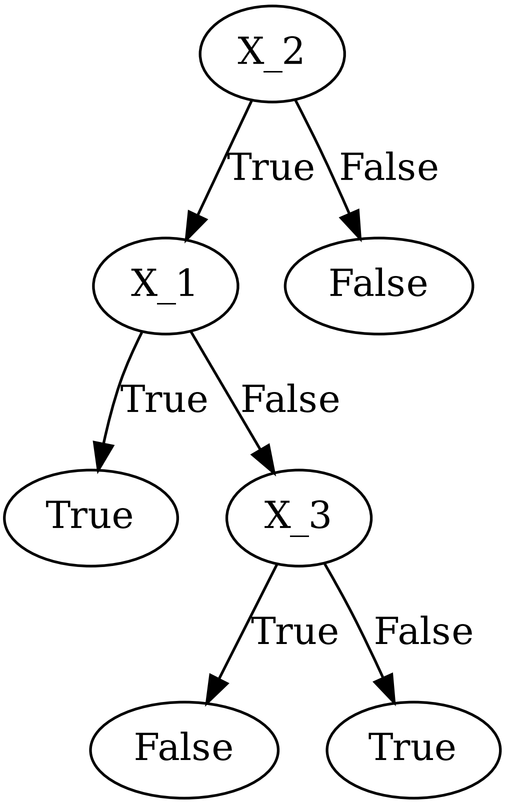
(c) $k - DNF$:
We make use of Theorem 1.3 (a).
We can map the binary tree t of height k to a boolean function $f : \{0,1\}^n \to \{0,1\}$ in $k - DNF$ by following algorithm:
For each path p in t, that leads to a `TRUE` classification, we can create the conjunction $c_p := \bigwedge_{l_i \in p} l_i$, where l_i is the literal of the variable x_i on the path. Since t is of height k, p at most contains k literals, thus c_p at most contains k literals. We can now define f in $k - DNF$ as follows:

$$f := \bigvee_{p \in t} c_p = \bigvee_{p \in t} \bigwedge_{l_i \in p} l_i$$

$k - CNF$:
We use the above algorithm for $k-DNF$ to create a boolean function $f' : \{0,1\}^n \to \{0,1\}$ in $k - DNF$, however, for this, we do only use the paths p in t, that lead to a `FALSE` classification. Therefore, f' is then `TRUE`, iff the decision tree t classifies `FALSE`.
By negating f', we can get our desired function f in $k - CNF$:

$$f := \neg f' = \neg \bigvee_{p \in t} \bigwedge_{l_i \in p} l_i = \bigwedge_{p \in t} \bigvee_{l_i \in p} \neg l_i$$

Exercise 4



Reasoning (see Appendix for code):

Feature Set: $[X_1, X_2, X_3]$
Gains for each feature $[(X_2, 0.5487949406953987), (X_1, 0.04879494069539858), (X_3, 0.04879494069539858)]$
Splitting using feature X_2

Feature Set: $[X_1, X_3]$
Gains for each feature $[(X_1, 0.31127812445913283), (X_3, 0.31127812445913283)]$
Splitting using feature X_1

Feature Set: $[X_3]$
Gains for each feature $[(X_3, 1.0)]$
Splitting using feature X_3

Exercise 5

(a) $x_0' = [0,0,0,1,0]^\top$, since $\langle a, x\rangle = \frac{1}{3}\cdot(3\cdot 0 + 1\cdot 0 + 2\cdot 0 + -2\cdot 1 + 3\cdot 0) = -\frac{2}{3} = b$

(b) $\langle a, x_1\rangle + \frac{2}{3} = \frac{1}{3}\cdot(3+1+2-2+3) + \frac{2}{3} = \frac{7+2}{3} = 3 \geq 0$
$\langle a, x_1'\rangle + b = \frac{1}{3}\cdot(3-1-2+2-3) + b = \frac{-1+2}{3} = \frac{1}{3} \geq 0$
Therefore, x_1, x_1' are contained in the same halfspace.

(c) Let $X \in P$ be a point on the hyperplane. Then the length of the projection of $(x_1 - X)$ is $\frac{|\langle a, x_0 - X\rangle|}{\|a\|} = \frac{|\langle a, x_0\rangle - \langle a, X\rangle|}{\|a\|} = \frac{\langle a, x_1\rangle + b}{\|a\|} = \frac{3}{26+\frac{1}{9}} = \frac{27}{235} \simeq 0.11489$

Exercise 6

Since we are in $\mathbb{R}^3$ we can divide the grape by all 3 dimensions twice, thus splitting the grape with 3 strikes into $2^3 = 8$ pieces. The last strike cannot split the grape by another dimension. We can only split at most 7 pieces in half, thus a maximum of 15 pieces. The swordmasters claims are invalid.

Appendix

Code for Exercise 2

```python
from math import sqrt
#from statistics import mode

training_set = [
    ((-4, -2.1, -1), -1),
    ((-3.6, -1.4, 0.2), 1),
    ((1, -0.2, -0.3), 1),
    ((0.3, -0.5, -0.5), 1),
    ((-2, -3.5, -1), -1),
    ((-4.2, -4, 0.2), 1),
    ((-1.3, -0.1, -3), 1),
    ((-0.7, 0.9, -0.7), 1),
    ((1, 2, 1.4), 1),
    ((2.6, -1.5, 0.2), 1),
    ((2, 4.3, -0.7), -1),
    ((0.6, 0.4, 0.2), -1),
    ((2.9, -1.7, 3.6), -1),
    ((3.6, 0.4, -2.5), -1),
    ((1.2, 4, 1.2), -1),
    ((-1, 0.5, 0.5), -1),
    ((3, 2.7, 2.3), -1),
    ((4, -3, 2.2), -1),
    ((0.1, 0.1, 3.5), -1),
    ((2.8, 1.2, 2.4), -1)
]

test_set = [
    (1, -2, 0),
    (4, -0.5, 2),
    (1, 1.5, -2.5),
    (-2, -1, -2),
    (-4, -1, -1)
]


def manhattan_distance(x, y):
    assert len(x) == len(y)
    sum = 0
    for i in range(len(x)):
        sum += abs(x[i] - y[i])
    return sum

```

```python

def euclidean_distance(x, y):
    assert len(x) == len(y)
    sum = 0
    for i in range(len(x)):
        sum += pow(x[i] - y[i], 2)
    sum = sqrt(sum)
    return sum


def get_k_nearest_neighbors(training_set, test, k, distance_func):
    assert callable(distance_func)
    distances = list()
    for data in training_set:
        distance = distance_func(data[0], test)
        distances.append((data, distance))
    distances.sort(key=lambda x: x[1]) # sort by distance
    neighbors = list()
    for i in range(k): # get data of k nearest neighbors
        neighbors.append(distances[i][0])
    return neighbors


def predict_classificataion(training_set, test, k, distance_func):
    assert callable(distance_func)
    neighbors = get_k_nearest_neighbors(training_set, test, k, distance_func)
    classifications = [neighbor[1] for neighbor in neighbors] # list of all classifications
    #prediction = mode(classifications) # get classification most often in k nearest neighbors
    #return prediction
    count_neg = classifications.count(-1)
    count_pos = classifications.count(1)
    assert count_neg + count_pos == k
    if count_pos > count_neg:
        return 1
    if count_pos < count_neg:
        return -1
    if count_pos == count_neg:
        return 0


if __name__ == '__main__':
    print('Classification: k=2 Manhattan Distance')
    for test in test_set:
        prediction = predict_classificataion(
            training_set, test, 2, manhattan_distance)
        print(f'Test {test}: Prediction {prediction}')
    print('\n')
    print('Classification: k=3 Manhattan Distance')
    for test in test_set:
        prediction = predict_classificataion(
            training_set, test, 3, manhattan_distance)
        print(f'Test {test}: Prediction {prediction}')
    print('\n')
    print('Classification: k=2 Euclidean Distance')
    for test in test_set:
        prediction = predict_classificataion(
            training_set, test, 2, euclidean_distance)
```

```
        print(f'Test {test}: Prediction {prediction}')
    print('\n')
    print('Classification: k=3 Euclidean Distance')
    for test in test_set:
        prediction = predict_classificataion(
            training_set, test, 3, euclidean_distance)
        print(f'Test {test}: Prediction {prediction}')
```

Code for Exercise 4

```
from math import log2
from graphviz import Digraph
import queue

feature_set = [0, 1, 2]

example_set = [
    ((False, False, False), False),
    ((False, False, True), False),
    ((False, True, False), True),
    ((False, True, True), False),
    ((True, False, False), False),
    ((True, False, True), False),
    ((True, True, False), True),
    ((True, True, True), True),
]


class Node:
    def __init__(self, value):
        self.left = None
        self.left_label = True
        self.value = value
        self.right = None
        self.right_label = False


def count_pos(example_set) -> int:
    classifications = [example[1] for example in example_set]
    return classifications.count(True)


def count_neg(example_set) -> int:
    classifications = [example[1] for example in example_set]
    return classifications.count(False)


def partition_example_set(feature, example_set) -> tuple:
    list_true = list()
    list_false = list()
    for example in example_set:
        if example[0][feature]:
            list_true.append(example)
        else:
            list_false.append(example)
    return (list_true, list_false)


def entropy(example_set) -> float:
    pos = count_pos(example_set)
    neg = count_neg(example_set)
```

```python
    assert pos + neg == len(example_set)
    pos_frac = pos / float(pos + neg)
    neg_frac = 1 - pos_frac
    if pos_frac == 0:
        return -neg_frac * log2(neg_frac)
    if neg_frac == 0:
        return -pos_frac * log2(pos_frac)
    result = -(pos_frac * log2(pos_frac) + neg_frac * log2(neg_frac))
    assert 0 <= result and result <= 1
    return result


def remainder(feature, example_set) -> float:
    (list_true, list_false) = partition_example_set(feature, example_set)
    true_frac = len(list_true) / float(len(example_set))
    false_frac = 1 - true_frac
    result = true_frac * entropy(list_true) + false_frac * entropy(list_false)
    return result


def gain(feature, example_set) -> float:
    result = entropy(example_set) - remainder(feature, example_set)
    return result


def decision_tree(feature_set, example_set) -> Node:
    if len(example_set) == 0:
        return Node(False)  # arbitrary value

    if count_pos(example_set) == len(example_set):
        return Node(True)
    if count_neg(example_set) == len(example_set):
        return Node(False)

    gains = [(feature, gain(feature, example_set)) for feature in feature_set]
    gains.sort(key=lambda x: x[1], reverse=True)
    max_gain = gains[0]
    partition_feature = max_gain[0]
    (true_frac, false_frac) = partition_example_set(
        partition_feature, example_set)
    new_feature_set = list(feature_set)
    new_feature_set.remove(partition_feature)

    print(f'Feature Set: {list(map(lambda feature: f"X_{feature+1}", feature_set))}')
    print(f'Gains for each feature {list(map(lambda x: (f"X_{x[0]+1}", f"Gain: {x[1]}"), gains))}')
    print(f'Splitting using feature X_{partition_feature + 1}')

    node = Node(f'X_{partition_feature + 1}')
    node.left = decision_tree(new_feature_set, true_frac)
    node.left_label = True
    node.right = decision_tree(new_feature_set, false_frac)
    node.right_label = False

    return node
```

```python
def draw_decision_tree(tree):
    dot = Digraph('Decision-Tree')
    q = queue.Queue()
    q.put(('root', tree))
    while not q.empty():
        (node_name, node) = q.get()
        dot.node(name=node_name, label=str(node.value))
        if node.left:
            child_name = node_name+str(node.value) + \
                str(node.left_label)+str(node.left.value)
            dot.node(name=child_name, label=str(node.left.value))
            dot.edge(node_name, child_name, label=str(node.left_label))
            q.put((child_name, node.left))

        if node.right:
            child_name = node_name + \
                str(node.value)+str(node.right_label)+str(node.right.value)
            dot.node(name=child_name, label=str(node.right.value))
            dot.edge(node_name, child_name, label=str(node.right_label))
            q.put((child_name, node.right))
    dot.render()


if __name__ == '__main__':
    tree = decision_tree(feature_set, example_set)
    draw_decision_tree(tree)
```