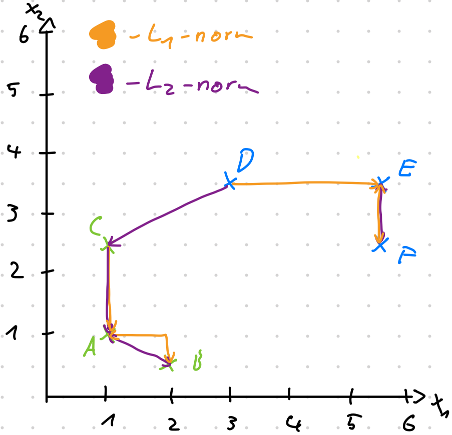
Problem 1



ⓐ

L_1-norm : $d_{L_1}(x_i, x_j) = \sum_{d=1}^{D} |x_{id} - x_{jd}|$

A: $d_{L_1}(A,B) = 1 + 0{,}5 = 1{,}5$

B: $d_{L_1}(B,A) = d_{L_1}(A,B) = 1{,}5$

C: $d_{L_1}(C,A) = 0 + 1{,}5 = 1{,}5$

D: $d_{L_1}(D,C) = 2 + 1 = 3$

$d_{L_1}(D,E) = 2{,}5 + 0 = 2{,}5$

$d_{L_1}(D,E) < d_{L_1}(D,C) \Leftrightarrow$ D's nearest neighbour is E (with L_1-norm)

E: $d_{L_1}(E,F) = 0 + 1 = 1$

F: $d_{L_1}(F,E) = d_{L_1}(E,F) = 1$

ⓑ

$d_{L_2}(x_i, x_j) = \sqrt{\sum_{d=1}^{D} (x_{id} - x_{jd})^2}$

A: $d_{L_2}(A,B) = \sqrt{1 + 0{,}5^2} = \sqrt{1{,}25} = \frac{\sqrt{5}}{2} = \frac{1}{2} \cdot \sqrt{5}$

B: $d_{L_2}(B,A) = d_{L_2}(A,B) = \frac{1}{2} \cdot \sqrt{5}$

C: $d_{L_2}(C,A) = \sqrt{0 + 1{,}5^2} = 1{,}5$

D: $d_{L_2}(D,C) = \sqrt{4+1} = \sqrt{5} \approx 2{,}2$

$d_{L_2}(D,E) = \sqrt{2{,}5^2 + 0} = 2{,}5$

$d_{L_2}(D,C) < d_{L_2}(D,E) \Leftrightarrow$ D's nearest neighbour is C (with L_2-norm)

E: $d_{L_2}(E,F) = \sqrt{0 + 1^2} = 1$

F: $d_{L_2}(F,E) = d_{L_2}(E,F) = 1$

ⓒ

The classification is correct for the L_1-norm.

For L_2-norm it is incorrect and D should have class 1 instead of 2.

Problem 2

(a)

$N = 16 + 32 + 64 = 112$

$P(class(x_{new}) = A) = \frac{16}{N} = \frac{16}{112} = \frac{4}{28} = \frac{1}{7}$

$P(class(x_{new}) = B) = \frac{32}{112} = \frac{2}{7}$

$P(class(x_{new}) = C) = \frac{64}{112} = \frac{4}{7}$

The likelihood of the new point x_{new} being from class A is $\frac{1}{7}$, being from class $B = \frac{2}{7}$ and being from class $C = \frac{4}{7}$.

(b)

In case of weighted (by distance) version of k-NN it is impossible to predict a new class due to the fact that it is unknown where the points are located.

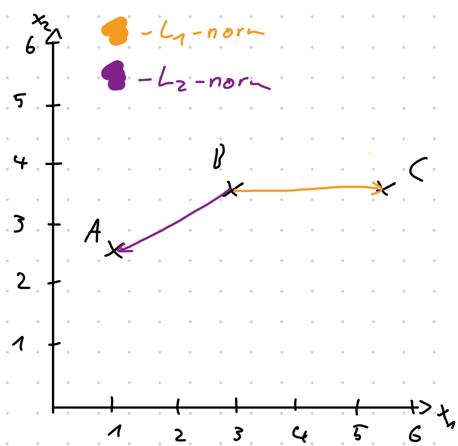
Problem 5

Disprove: by counterexample

$A(1/2.5)$

$B(3/3.5)$

$C(5.5/3.5)$



$d_{L1}(B, A) = 2 + 1 = 3$

$d_{L1}(B, C) = 0 + 2{,}5 = 2{,}5$

↳ The nearest neighbor from B with L_1-norm is C because $d_{L1}(B, A) > d_{L1}(B, C)$.

$d_{L2}(B, A) = \sqrt{4 + 1} = \sqrt{5} \approx 2{,}23$

$d_{L2}(B, C) = \sqrt{0 + 2{,}5^2} = 2{,}5$

$d_{L2}(B, A) < d_{L2}(B, C)$

↳ The nearest neighbor from B with L_2-norm is A. ↯

Problem 6

(a) There exist a decision tree for the data but not with depth 1. To build the decision tree one has to split on one of the feature and there is no possible split to achieve a decision tree with depth 1.

(b) It is possible to transform the feature space so that the diagonal is horizontal or vertical. Then the decision tree of depth 1 has 100 % accuracy.

Problem 7

(a)
$$i_H(y) = -\sum_{c_i \in C} \pi_{c_i} \cdot \log(\pi_{c_i}) \qquad \pi_c = p(x=c \mid y)$$

$$\pi_{x_1} = p(x_1 = c \mid y) = p(x_1 = T \mid y) + p(x_1 = I \mid y)$$

$$p(x_1 = T \mid y) = p(x_1 = T \mid y = W) + p(x_1 = T \mid y = L)$$

$$p(x_1 = T \mid y = W) = \frac{2}{4}$$

$$p(x_1 = T \mid y = L) = \frac{3}{6}$$

$$= \frac{2}{4} + \frac{3}{6} = 1$$

$$p(x_1 = I \mid y) = 1 \quad \text{// same distribution like } p(x_1 = T \mid y)$$

$$= 1 + 1 = 2$$

$$\pi_{x_2} = p(x_2 = c \mid y) = p(x_2 = M \mid y) + p(x_2 = P \mid y)$$
$$= \left(\frac{2}{4} + \frac{2}{6}\right) + \left(\frac{2}{4} + \frac{4}{6}\right) = \frac{5}{6} + \frac{7}{6} = \frac{12}{6} = 2$$

$$\pi_{x_3} = p(x_3 = c \mid y) = p(x_3 = S \mid y) + p(x_3 = C \mid y)$$
$$= \left(\frac{3}{4} + \frac{2}{6}\right) + \left(\frac{1}{4} + \frac{4}{6}\right) = \frac{13}{12} + \frac{11}{12} = 2$$

$$i_H(y) = -3(2 \cdot \log(2)) = -6 \cdot \log(2) \approx -6 \cdot 0{,}301$$
$$\approx -1{,}806$$

Problem 3: I think that we encounter a scaling issue regardin the acceleration and the maximal velocity. Usually, the acceleration is specified in the unit m/s^2. Altough it is not clearly stated in the shown data set, I assume the acceleration is specified in m/s^2 in this case. To prevent this we could use standardization. A decision tree would perform better because it does not rely on distance measurements. Further, a decision tree always looks only on a single attribute and does not capture relations between several attributes.

Additionally, we don't know much about the size of the data set. It could be the case, that we have significantly not enough data, as it is only secured we have 100 samples per class. The only way to solve this problem is to gather more data. A decision tree would not help with this issue.

In terms of the quality of the data, it can be stated that it has some flaws. Taking a look at the attribute 'weight' we can observe that the van is lighter than the car. This seems somewhat strange. Furthermore, the van has the fastest acceleration under the assumption that it is measured in m/s^2. To solve this problem we could get rid of erroneous data or if possible, recollect it. A decision tree would have the same problem.