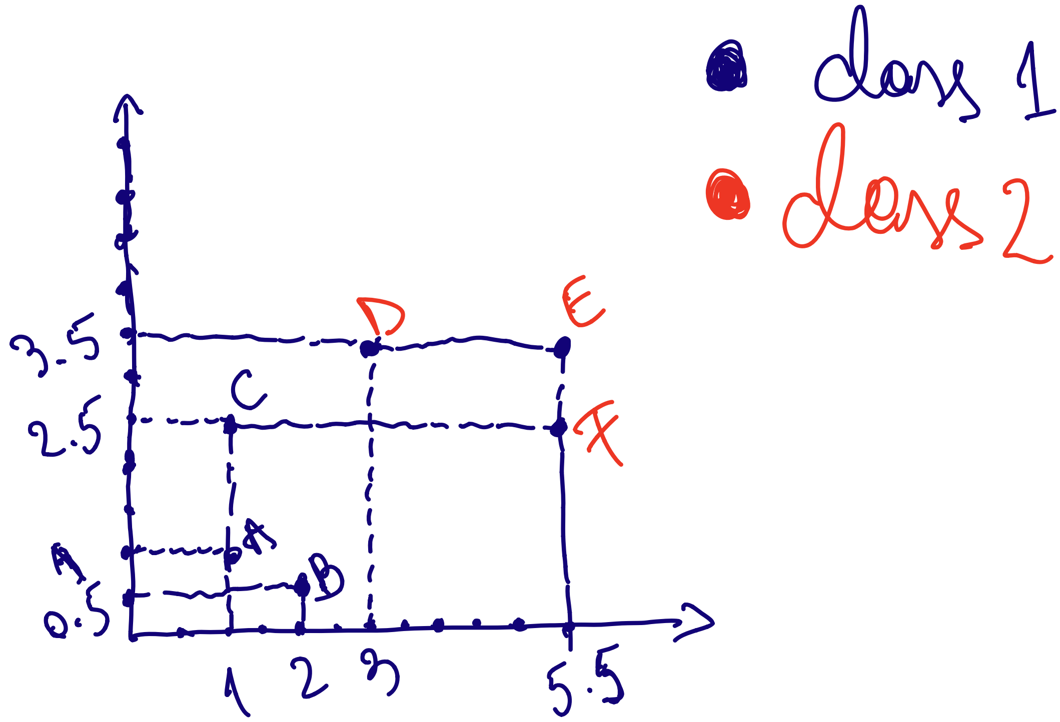
NAME: THEODOR STOICAN

ID: ge93pas

MATRICULATION NUMBER: 03725732

ASSIGNMENT 2

Problem 1



a) Point A:

According to the L-1 norm, either B or C are the nearest neighbours for A. In this case, by convention, we choose B.

$$d(A,B) = |2-1| + |0.5-1| = 1.5$$

(class 1)

Point B:

Similarly, $d(B,A) = |1-2| + |1-0.5|$
$= 1.5$ (class 1)

Point C:

$d(C,A) = |2.5-1| + |1-1| = 1.5$
(class 1)

Point D:

$d(D,E) = |5.5-3| + |3.5-3.5| = 2.5$
(class 2)

Point E:

$d(E,F) = |3.5-2.5| + |5.5-5.5|$
$= 1$ (class 2)

Point F:

$d(F,E) = |2.5-3.5| + |5.5-5.5|$
$= 1$ (class 2)

b) Point A:

$$d(A,B) = \sqrt{(2-1)^2 + (1-0.5)^2}$$

$$= \sqrt{1 + (0.5)^2}$$

$$= \sqrt{1 + \frac{1}{4}} = \sqrt{\frac{5}{4}} = \frac{\sqrt{5}}{2}$$

(class 1)

Point B:

$$d(B,A) = \frac{\sqrt{5}}{2}$$ (as before)

(class 1)

Point C:

$$d(C,A) = \sqrt{(2.5-1)^2 + (1-1)^2}$$

$$= \sqrt{(1.5)^2} = 1.5$$

(class 1)

Point D:

$$d(D,C) = \sqrt{(3-1)^2 + (3.5-2.5)^2}$$

$$= \sqrt{4+1} = \sqrt{5}$$

(class 1)

Point E:

$$d(E,F) = \sqrt{(3.5-2.5)^2}$$
$$= \sqrt{1} = 1 \text{ (class 2)}$$

Point F:

$d(F,E) = 1$ (as before) (class 2)

c) As we can see from the previous computations, using the L1 norm yields better performance since it classifies correctly all the points, whereas using the L_2 norm leads to a misclassification (point Δ).

Problem 2

a) The probability that x_{new} is assigned to a class X is the following:

$$P(y = X \mid x_{new}, K = N_A + N_B + N_C) =$$

$$= \frac{1}{N_A + N_B + N_C} \sum_{i \in N_K(x_{new})} \mathbb{I}(y_i = X)$$

This value will reach its maximum for class X = C since the instances that belong to class C are the most numerous ones.

$\left(\frac{N_C}{N_A + N_B + N_C} \right.$ is the longest value $\left. \right)$

Therefore, a new point will always be assigned to class C.

b) In this case, the classification score will improve, since new

points which are within one particular cluster and significantly farther from the other clusters will be assigned the same class as the particular cluster they are part of (given that the other clusters are sufficiently "far away" so that their score may not be greater than the current score).
Long story short, the answer is dependent on the positions of the clusters and the new point!

Problem 3

The diagonal line has equation $x_1 - x_2 = 0$

Since, due to practical reasons, decision trees rely on horizontal and vertical splits only, no matter what we do, we cannot approximate the diagonal with a single such line (depth=1). Therefore, we cannot have 100% accuracy on this dataset with a decision tree of depth 1.

Problem 4

a) $i_H(y) = \sum_c -p(y=c)\log p(y=c)$

$p(y = \text{Win}) = \frac{4}{10} = \frac{2}{5}$

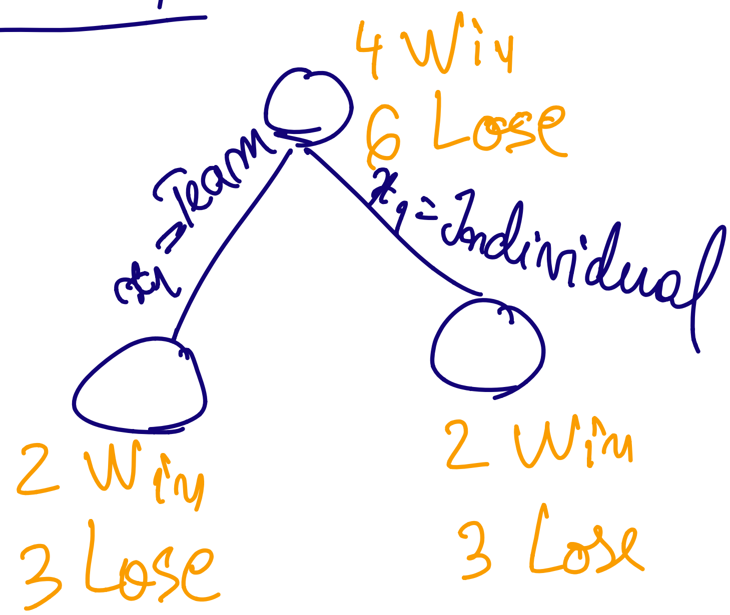
$p(y = \text{Lose}) = 1 - \frac{2}{5} = \frac{3}{5}$

$\Rightarrow i_H(y) = -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) \approx 0.52 + 0.44$

≈ 0.96

b) We have 3 features and therefore 3 possible splits that we can make.

Split on x_1



$$i_H(t_L) = -p(y=\text{Win})\log p(y=\text{Win}) - p(y=\text{Lose})\log p(y=\text{Lose})$$

$$= -\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5}$$

$$\approx 0.96$$

$i_H(t_R) \approx 0.96$ (same as $i_H(t_L)$ because of the distribution of classes)

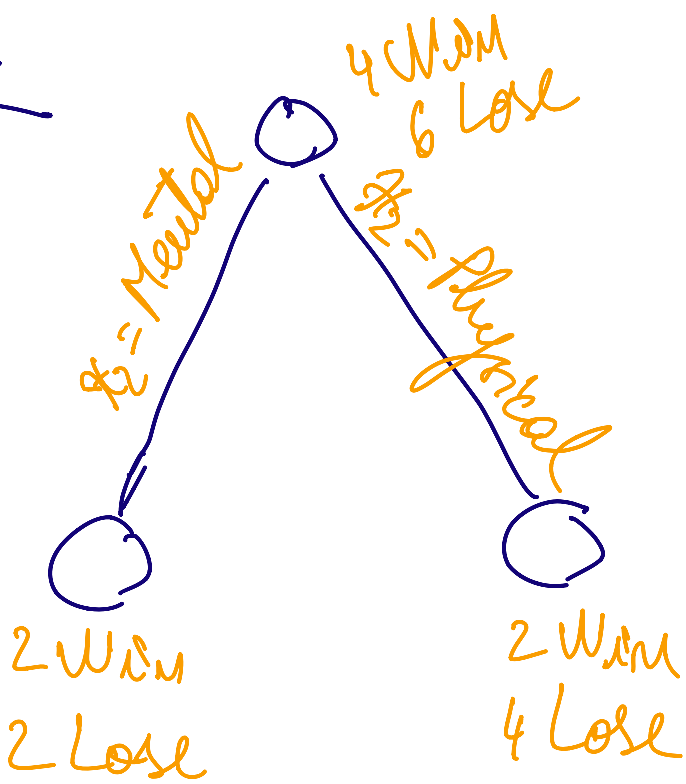
Hence, $\Delta i_H(x_1=\text{Team}, t) = i_H(t) - \frac{5}{10} i_H(t_L) - \frac{5}{10} i_H(t_R)$

$$\Rightarrow \Delta i_H(x_1=\text{Team}, t) = 0.96 - \frac{1}{2}(0.96+0.96)$$

$$= 0.96 - \frac{1}{\cancel{2}} \cdot \cancel{2} \cdot 0.96$$

$$= \boxed{0}$$

Split on x_2



$$i_H(t_L) = -P(y=\text{Win}) \log P(y=\text{Win}) - P(y=\text{Lose}) \log P(y=\text{Lose}) =$$

$$= -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}$$

$$= -\log\frac{1}{2} = -\log 2^{-1} = 1$$

$$i_H(t_R) = -\frac{2}{6}\log\frac{2}{6} - \frac{4}{6}\log\frac{4}{6}$$

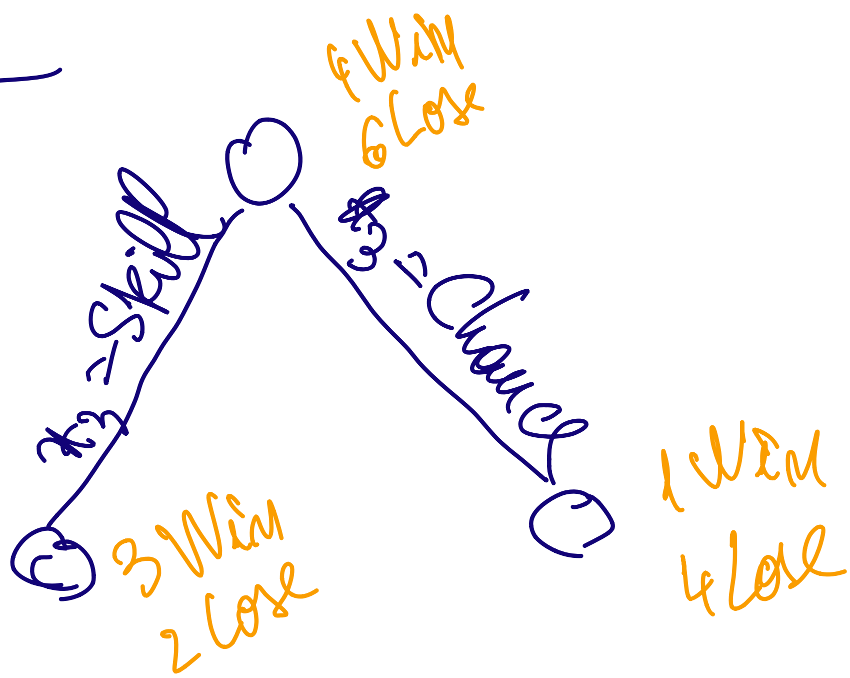
$$= -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}$$

$$\approx 0.91$$

$$\Rightarrow \Delta i_H(x_2 = \text{Mental}, t) = 0.96 - \frac{4}{10}\cdot 1 - \frac{6}{10}\cdot 0.91$$

$$\Rightarrow \Delta i_H(x_2 = \text{Mental}, t) \approx 0.014$$

Split on x_3



$$i_H(t_L) = -\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5}$$
$$\approx 0.96$$

$$i_H(t_R) = -\frac{1}{5}\log\frac{1}{5} - \frac{4}{5}\log\frac{4}{5}$$
$$\approx 0.72$$

$$\Delta i_H(x_3 = \text{Skill}, t) = 0.96 - \frac{5}{10}\cdot 0.96 - \frac{5}{10}\cdot 0.72$$
$$\approx \boxed{0.12}$$

Given all of these results, the longest Δi_H is 0.12 and the optimal decision tree of depth $=1$ can be obtained by splitting on x_3.