LEIC-T 2024/2025
Aprendizagem - Machine Learning
Homework I
Deadline 27/9/2024 21:00
Submit on Fenix as pdf

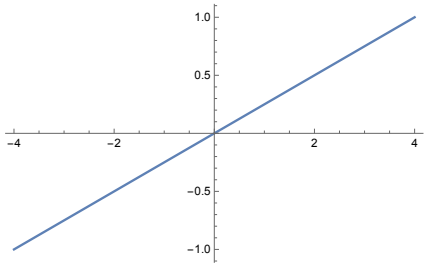
I) Correlation (3 pts)

Compute the correlation (Pearson correlation) and **Spearman's rank** for two variables $\mathbf{x}_1$ and $\mathbf{x}_2$=f($\mathbf{x}_1$). Are the values the same or different, please justify. (Indicate all computational steps!)

(a) (1 pts)

$\mathbf{x}_1$ = (-4, -2, 0, 2, 4)

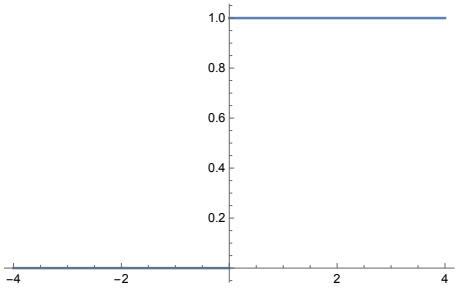
for linear function



f($\mathbf{x}_1$)=0.25* $\mathbf{x}_1$
$\mathbf{x}_2$=f($\mathbf{x}_1$).

(b) (1pts)
$\mathbf{x}_1$ = (-4, -2, 0, 2, 4)

for $sign_0$ also called unit step function

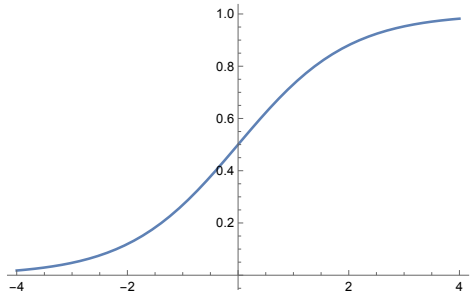


f($\mathbf{x}_1$)=$\begin{cases} 1 & x_1 \geq 0 \\ 0 & x_1 < 0 \end{cases}$
$\mathbf{x}_2$=f($\mathbf{x}_1$).

LEIC-T 2024/2025
Aprendizagem - Machine Learning
Homework I
Deadline 27/9/2024 21:00
Submit on Fenix as pdf

(c) (1pts)
$\mathbf{x}_1$ = (-4, -2, 0, 2, 4)

for sigmoid also called logistic function



$f(\mathbf{x}_1)=\frac{1}{1+e^{-\mathbf{x}1}}$ =1/(1+Exp(-$\mathbf{x}_1$)
$\mathbf{x}_2$=f($\mathbf{x}_1$).

II) Decision Trees (5 pts)

Training Set:

Weekend	*Weather*	*Tired*	*What to Do?*
Yes	*Sunny*	*No*	*Go for a walk*
No	*Rain*	*Yes*	*TV*
No	*Rain*	*No*	*Reading*
No	*Claudy*	*No*	*Go for a walk*
Yes	*Claudy*	*Yes*	*Reading*

(a) (2 pts)
Determine the root of decision tree using the ID3 algorithm with the target "*What to Do?*".
Indicate the calculation. (Indicate all computational steps!)

(b) (2 pts)
Determine the decision tree using the ID3 algorithm with the target "*What to Do?*"
Draw your decision tree, if there is a tie chose the attribute *Weekend* before the attribute *Tired*.
(YOU DO NOT NEED TO INDICATE THE CALCULATION STEPS)

TÉCNICO LISBOA

LEIC-T 2024/2025
Aprendizagem - Machine Learning
Homework I
Deadline 27/9/2024 21:00
Submit on Fenix as pdf

(c) (1 pts)
Draw confusion matrix for the test set

Test Set:

Weekend	*Weather*	*Tired*	*What to Do?*
Yes	*Sunny*	*No*	*Reading*
No	*Rain*	*Yes*	*TV*
No	*Sunny*	*No*	*Reading*
No	*Claudy*	*Yes*	*Go for a walk*
Yes	*Rain*	*Yes*	*Reading*

using the learnt decision tree form b)

III Software Experiments (2pts)

Download the jupyter notbook HM1_DT_24.ipynb.

(a) (1pts)
Split the data using the command (in the notebook)

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=**value**,stratify=y,random_state=**your_group number**)

Partition data with the *train_test*_split values *0.1, 0.9* and indicate the depth and the accuracy of each the decision tree (if you have no group yet, put the last three digits of student nr).

Why is the depth of the decision tree bigger and the accuracy higher for the *train_test_split* values *0.9* than for the value *0.1*?

(b) (1pts)
Now perform the same experiment with the *train_test_split* value 0.9 without the command *stratify=y*, see

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=**0.9**, ,random_state=**your_group number**)

Why is now the accuracy lower? Please write only one/two sentence/s… Hint, look in the scikit-learn documentation for meaning of the command *stratify=y.*