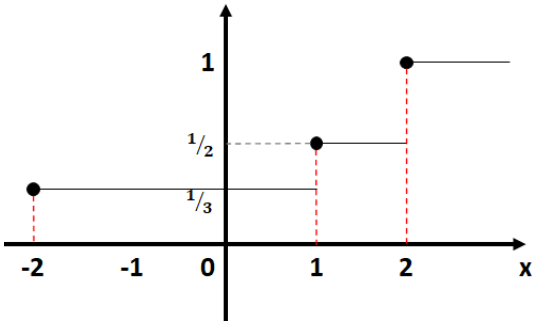
COMP 5450 Machine Learning

**** Unless specified, you can choose to code your results or calculate by hand. If coded, you are required to submit your code (R or Python), include any libraries you use. If by hand, all steps of calculation should be shown. Any answer without rationalization will not be given points. Type your work when possible, unintelligible work will not be graded.***

Homework 1

1. The Cumulative Distribution Function $F(x)$ for some discrete random variable is given in the figure below. Please identify the corresponding Probability Mass Function $f(x)$. (**5pts**)



2. A coin we know nothing about has two sides {Head, Tail}, the coin is flipped 10 times, we observed 6 heads. The probability that the coin show heads is one of the following possibilities: $\frac{1}{3}$, $\frac{1}{2}$, or $\frac{2}{3}$. Which is more probable? (Hint: compare $Pr(H|p)$) (**10pts**)
3. We have four hypotheses $\{h_1, h_2, h_3, h_4\}$. The posterior probabilities for the hypotheses are $P(h_1|D) = 0.25$, $P(h_2|D) = 0.3$, $P(h_3|D) = 0.4$, and $P(h_4|D) = 0.05$ respectively. The set of possible classification of the new instance is $V = \{+, -\}$. We also have: $P(-|h_1) = 0$, $P(+|h_1) = 1$, $P(-|h_2) = 1$, $P(+|h_2) = 0$, $P(-|h_3) = 1$, $P(+|h_3) = 0$, $P(-|h_4) = 1$, $P(+|h_4) = 0$. What is the result from the Bayes optimal classifier? (**10pts**)
4. Using the Boston housing dataset (you can find the dataset on Blackboard). Implement linear regression on the training dataset, and report the intercept, slope, and R^2 values. Use these values to discuss the quality of the model. (**15pts**)

5. The hypothesis space is defined in the following table, each hypothesis is represented as a pair of 4-tuples. Please use the Naïve Bayes classifier to predict the target value *Run* for the instances: (**15pts**)
   i) <Sunny, Mild, Normal, Weak>
   ii) <Rain, Cool, High, Strong>

Day	Outlook	Temperature	Humidity	Wind	Run
D1	Sunny	Hot	Normal	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	Normal	Strong	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Mild	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Rain	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Strong	Yes
D10	Rain	Hot	Normal	Strong	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Sunny	Mild	Normal	Weak	No
D13	Overcast	Hot	Normal	Weak	Yes
D14	Sunny	Cool	High	Weak	No

6. Consider the hypothesis space defined over instances shown below, we characterize each hypothesis (apple taste) by 4-tuples. Please hand trace the CART classifier to build a decision tree, then predict the target value Taste=Sweet/Tart for the following instances:
   a) <Red, High, Some, No>
   b) <Red, Low, Some, Yes>
   c) <Yellow, Low, Some, No>
   d) <Green, High, None, No>
   e) <Green, Mid, Some, Yes>
   Now suppose the actual taste of the five apples above are actually "Sweet, Sweet, Sweet, Tart, Tart", what is the accuracy of the decision tree? Please show all the steps and include the corresponding confusion matrix for accuracy calculation. (**20pts**)

Melon	Color	Crispiness	Spot	Fragrant	Taste
1	Red	High	None	Yes	Sweet
2	Red	High	None	No	Sweet
3	Yellow	High	None	Yes	Sweet
4	Yellow	Mid	Some	No	Tart
5	Yellow	Low	None	No	Sweet
6	Yellow	Mid	Some	Yes	Tart
7	Green	Low	Some	No	Sweet
8	Red	Low	None	Yes	Tart
9	Red	Low	None	No	Sweet
10	Yellow	Mid	None	Yes	Sweet

7. In a simple neural network shown below, the corresponding weights and biases are given. Please discuss in your own words the general function of a neural network, and the advantages/disadvantages of a deeper network structure compared to a shallow network structure. **(5pts)**
   An unknown bias value b_3 is initialized as 0. Use backpropagation to update b_3 twice. Include all the calculation steps. Use the $learning\_rate = 0.1$. **(20pts)**

Parameters	Value
w_1	3.3
b_1	-1.4
w_2	-3.5
b_2	0.5
w_3	-1.2
w_4	-2.3
b_3	0

Train Input	Train Output
0	0
0.5	1
1	4