Problem Set 9

Problem 1: Classification learning using 1R
1-1)
```
Gender:
    Male → No (4 out of 8)
    Female → Yes (4 out of 7)

Income:
    20-30K → No (2 out of 4)
    30-40K → Yes (3 out of 5)
    40-50K → No (2 out of 4)
    50-60K → Yes (1 out of 2)

LifeIns:
    Yes → Yes (6 out of 9)
    No → No (4 out of 6)

Magazine:
    Yes → Yes (5 out of 8)
    No → No (4 out of 7)
```

1-2)
Gender: Overall Accuracy = (4 + 4) / 15 = 53.3%
Income: Overall Accuracy = (2 + 3 + 2 + 1) / 15 = 53.3%
LifeIns: Overall Accuracy = (6 + 4) / 15 = 66.7%
Magazine: Overall Accuracy = (5 + 4) / 15 = 60%
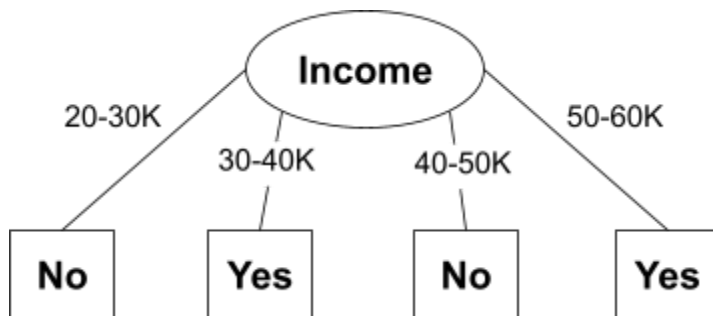Because the rules based on LifeIns have the highest overall accuracy of 66.7%, 1R selects them for the model.

1-3)

Actual	Predicted	
	Yes	**No**
Yes	6	2
No	3	4

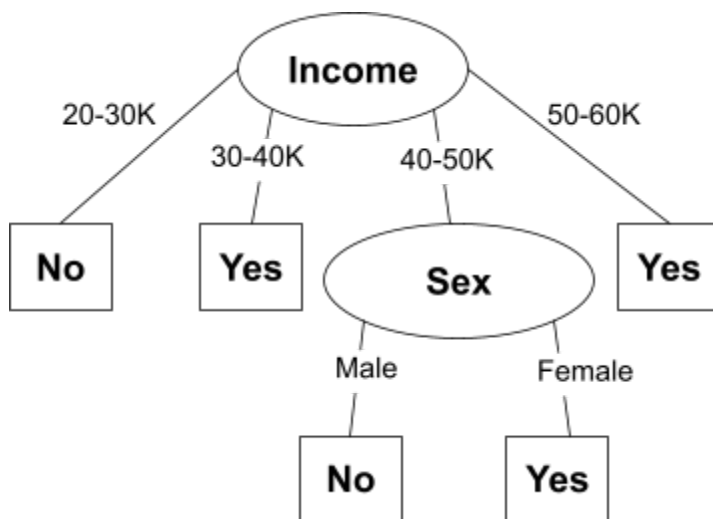Overall Error Rate = (3 + 2) / 15 = 33.3%

Problem 2: Learning a decision tree
2-1)

2-2)   accuracy = (4 + 5 + 3 + 2) / 15 = 93.3%
            goodness = 93.3 / 1 = 93.3

2-3)

2-4)
```
if Income == 20-30K:
     Magazine = No
elif Income == 30-40K:
     Magazine = Yes
elif Income == 40-50K and Sex == Male:
     Magazine = No
elif Income == 40-50K and Sex == Female:
     Magazine = Yes
elif Income == 50-60K:
     Magazine = Yes
```

Problem 3: Finding associations
3-1)
```
support = 4
confidence = 4/5
```

3-2)
```
Sex = Female, LifeIns = Yes, Watch = Yes
Sex = Feale, LifeIns = Yes, Magazine = Yes
Sex = Female, Watch = Yes, Magazine = Yes
```

3-3)
```
if LifeIns = Yes and Watch = Yes
then Magazine = Yes
if LifeIns = Yes and Magazine = Yes
then Watch = Yes
```

3-4)
```
Confidence of the first one = 5/6 < 0.9
Confidence of the second one = 5/7 < 0.9
Therefore, none of the two rules will be selected by the algorithm.
```

Problem 4: Discretizing numeric attributes
4-1)
```
Width = (56-20)/3 = 12
Bins: (0, 32], (32, 44], (44, infinity)
```

4-2)
```
(38+39)/2 = 38.5
(42+43)/2 = 42.5
Bins: (0,38.5], (38.5, 42.5], (42.5, infinity)
```