Explainable AI (CSE615)

Quiz-3, Date: April 23, 2025

Max Time: 30 mins | Winter 2025 | Max marks: 20

1. Express the following concepts in *ALC* description logic using the atomic concepts *Animal*, and *Fish*, and the roles *Tail*, *Leg*, and *Eat*. [4]

   (a) An animal that has a tail.
   Soln: $Animal \sqcap \exists Tail.\top$.

   (b) An animal that has a tail and four legs.
   Soln: Not expressible in ALC.

   (c) An animal that eats only fish.
   Soln $Animal \sqcap \forall Eat.Fish$

   (d) An animal that eats only things that themselves eat only fish.
   Soln: $Animal \sqcap \forall Eat.\forall Eat.Fish$

2. Identify the RDF and RDFS statements among the following. [4]

   1. :MotorVehicle rdfs:subClassOf :Vehicle .
   2. <http://example.org/bob#me> <http://xmlns.com/foaf/0.1/knows> <http://example.org/alice#me> .
   3. :author rdfs:domain :Document .
   4. <http://example.org/bob#me> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person> .

   Soln:
   1. RDFS.
   2. RDF.
   3. RDFS.
   4. RDF.

3. State one similarity and a difference among the following pairs. [2]

   (a) URN and IRI.
   Soln. Sim: Both are identifiers that can uniquely and unambiguously identify resources.
   Diff: URN - ASCII encoding, IRI - Universal Character Set (UTF-8)

   (b) Blank Nodes and Literals.
   Soln. Sim: Both doesn't need an IRI.
   Diff: Blank Node - can have outgoing edge, Literals - No outgoing edges.

4. You are using a Random Forest model to predict whether a customer will move away (churn) from a subscription service. The model uses the following 4 features: [5]

   - Feature 1: Number of months as a customer (continuous)
   - Feature 2: Monthly spending (continuous)

- Feature 3: Number of customer service calls made (integer)
- Feature 4: Account type (binary: 0 = basic, 1 = premium)

TreeSHAP Data: You are provided with the following data for the Decision Trees in the Random Forest model, showing the splits and leaf values:

- Tree 1: Split 1: Feature 1 (Months as customer) $\leq 18 \rightarrow$ Left node (Leaf value: 0.2), Right node (Leaf value: 0.8)

  Split 2: Feature 2 (Monthly spending) $\leq Rs.50 \rightarrow$ Left node (Leaf value: 0.1), Right node (Leaf value: 0.7)
- Tree 2: Split 1: Feature 4 (Account type) $= 1 \rightarrow$ Left node (Leaf value: 0.9), Right node (Leaf value: 0.6)
- Tree 3: Split 1: Feature 3 (Number of customer service calls) $\leq 2 \rightarrow$ Left node (Leaf value: 0.3), Right node (Leaf value: 0.6)

The leaf value is the accuracy for all the data points reaching that leaf node.

The feature values for a particular customer are as follows.

- Feature 1 (Months as customer): 24 months
- Feature 2 (Monthly spending): Rs.100
- Feature 3 (Number of customer service calls): 3
- Feature 4 (Account type): 1 (premium account)

The model's prediction output for this customer is 0.733 (the Random Forest prediction is the average of all 3 trees), meaning a 73.3% probability of cancellation. Using the TreeSHAP method, how do you calculate and interpret the SHAP values for each feature in this prediction? Note that, the estimated base value is the average of all leaf values in the forest.

Solution: We approximate the expected base value ϕ_0 as the average of all leaf values in the forest.

- Tree 1 leaves: 0.1, 0.2, 0.7, 0.8 $\rightarrow$ mean = (0.1 + 0.2 + 0.7 + 0.8) / 4 = 0.45
- Tree 2 leaves: 0.6, 0.9 $\rightarrow$ mean = (0.6 + 0.9) / 2 = 0.75
- Tree 3 leaves: 0.3, 0.6 $\rightarrow$ mean = (0.3 + 0.6) / 2 = 0.45

$\phi_0 = \frac{0.45+0.75+0.45}{3} = 0.55$

So, the total contribution of features is: $Prediction - \phi_0 = 0.733 - 0.55 = 0.1833$

Tree - 1 Contribution - From base value (0.45) to leaf value (0.7) = 0.25 Used features: F1 (Months) and F2 (Spending) $\rightarrow$ Assign 50% to each:

F1: +0.125

F2: +0.125

Tree 2 Contribution - From base (0.75) to prediction (0.9) = +0.15

Used feature: F4 (Account type)

F4: +0.15

Tree 3 Contribution - From base (0.45) to prediction (0.6) = +0.15

Used feature: F3 (Service calls)

F3: +0.15

SHAP value for each feature = average of its contributions across all trees:

- F1 (Months): 0.125 / 3 = +0.0417
- F2 (Spending): 0.125 / 3 = +0.0417
- F3 (Service calls): 0.15 / 3 = +0.05
- F4 (Account type): 0.15 / 3 = +0.05

Total contribution: 0.0417 + 0.0417 + 0.05 + 0.05 = 0.1833

The TreeSHAP explanation of the model's prediction:

- Base expected value: 0.55
- Model prediction: 0.733

SHAP values (contributions):

- F1 (Months as customer): +0.0417
- F2 (Monthly spending): +0.0417
- F3 (Customer service calls): +0.05
- F4 (Account type): +0.05

5. Given the following triples, convert them to Turtle format using Prefix, predicate, and object lists. [5]

- :Alice rdf:type :Student
- :Bob rdf:type :Professor
- :CS101 rdf:type :GraduateCourse
- :CS101 :offeredBy :CSDept
- :Alice :enrolledIn :CS101
- :Bob :teaches :CS101
- :CSDept rdfs:label "Department of Computer Science"@en
- :Student rdfs:subClassOf :Person
- :Professor rdfs:subClassOf :Person
- :GraduateCourse rdfs:subClassOf :Course
- :enrolledIn rdfs:domain :Student
- :enrolledIn rdfs:range :Course
- :teaches rdfs:domain :Professor
- :teaches rdfs:range :Course

- :offeredBy rdfs:domain :Course
- :offeredBy rdfs:range :Department

Solution -

```
@prefix : <http://example.org/univ#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

### Class hierarchy
:Student         rdfs:subClassOf :Person .
:Professor       rdfs:subClassOf :Person .
:GraduateCourse  rdfs:subClassOf :Course .

### Property constraints
:enrolledIn rdfs:domain :Student ;
            rdfs:range  :Course .

:teaches    rdfs:domain :Professor ;
            rdfs:range  :Course .

:offeredBy  rdfs:domain :Course ;
            rdfs:range  :Department .

### Instances and facts
:Alice  rdf:type     :Student ;
        :enrolledIn  :CS101 .

:Bob    rdf:type     :Professor ;
        :teaches     :CS101 .

:CS101  rdf:type     :GraduateCourse ;
        :offeredBy   :CSDept .

:CSDept rdfs:label   ``Department of Computer Science"@en .
```