(O) Infer a Surprise (1/4) [5 Points]

A popular task in natural language processing is called natural language inference (NLI). This task involves training a model to take two sentences and label whether the first sentence *entails* the second sentence. (Sentence 1 is said to entail Sentence 2 if Sentence 2 is guaranteed to be true whenever Sentence 1 is true). Here are some examples of entailment and non-entailment:

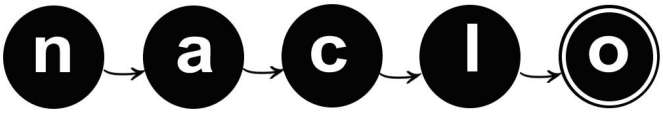
Sentence 1	**Sentence 2**	**Label**
The judge is 6 feet tall and the lawyer is 5 feet tall.	The judge is taller than the lawyer.	entailment
Lichen grows on every continent.	Lichen grows on Antarctica.	entailment
The dentist was born in Illinois.	The dentist was born in Chicago.	non-entailment
Lichen grows on every continent.	The judge is taller than the lawyer.	non-entailment
Lichen grows on every continent.	Lichen does not grow on every continent.	non-entailment

To get a computer to solve this task, the standard approach is to train the computer on many examples like the ones above. Ideally, the computer will solve the task by learning to understand the sentences and therefore figure out which sentences have meanings that entail the meanings of other sentences. However, sometimes a computational model will place too much weight on certain coincidences in the training data, and this tendency will cause it to make incorrect predictions. Suppose a model is trained on these examples:

Sentence 1	**Sentence 2**	**Label**
Rhode island is the smallest US state.	Rhode Island is smaller than Connecticut.	entailment
Rice is a type of grass.	Rice is a plant.	entailment
Ottawa is the capital city of Canada.	Ottawa is in Alberta.	non-entailment
No human has ever set foot on Mars.	Several animals have been to Mars.	non-entailment

A model trained on these examples might learn the following generalization:
If both sentences start with R, the sentence pair should be labeled entailment. Else, label non-entailment.
However, this conclusion is incorrect. It makes the wrong predictions for the following sentences:

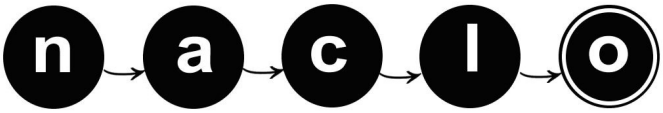
Sentence 1	**Sentence 2**	**Correct label**	**Prediction**
Regularization is useful.	Regularization is not useful.	non-entailment	entailment
No amphibians can echolocate.	Frogs cannot echolocate.	entailment	non-entailment



(O) Infer a Surprise (2/4)

A natural language inference model has been trained on the sentences in the following table (in practice, such a model would use a much larger training set, but we are displaying a small set to keep it manageable):

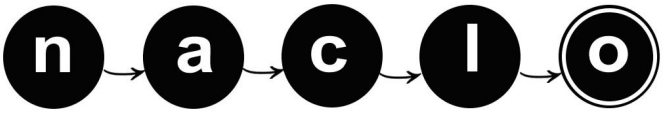
#	Sentence 1	Sentence 2	Label
1	the happiest walrus lives in Paris	the happiest walrus lives in Berlin	non-entailment
2	all dogs are mammals	all mammals are dogs	non-entailment
3	all dogs are mammals	no dog is not a mammal	entailment
4	we have stopped eating at Loretta's Diner	we once ate at Loretta's Diner	entailment
5	ice cream contains sugar	Chile is a narrow country	non-entailment
6	apples are red, and oranges are orange	oranges are orange	entailment
7	I love chocolate milk	I enjoy strawberry milk	non-entailment
8	the building is seventy feet tall	the building is over sixty-three feet in height	entailment
9	Charles Dickens was born in Portsmouth	the author of *A Tale of Two Cities* was born in a coastal city in the	entailment
10	ice cream contains sugar	ice cream has sugar in it	entailment
11	Mary knows that the vase broke	the vase broke	entailment
12	all mammals are welcome here	Paul the walrus is welcome here	entailment
13	the editor read the submission	the submission was read by the editor	entailment
14	the squirrel chased the chipmunk	the chipmunk chased the squirrel	non-entailment
15	the folder containing my passport is in the filing cabinet	my passport is in the filing cabinet	entailment
16	I have never seen a walrus	I have never seen a manatee	non-entailment
17	Augustus was the first emperor	Augustus was an emperor	entailment
18	every walrus enjoys swimming	most walruses enjoy yoga	non-entailment
19	etiquette demands that one display a certain degree of respect toward one's elders	nothing in life is ever free	non-entailment
20	I like Baltimore very much	I like Baltimore	entailment
21	Wilhelmina has a cousin	Wilhelmina has at least one aunt or uncle	entailment



(O) Infer a Surprise (3/4)

The model was then tested on many examples, and it got the following examples wrong:

#	Sentence 1	Sentence 2	Correct Label	Model Prediction
22	while the painter painted the furniture was covered with a plastic sheet	the painter painted the furniture	non-entailment	entailment
23	I have never, ever seen a walrus	I have never seen a walrus	entailment	non-entailment
24	the book on the table is blue	the table is blue	non-entailment	entailment
25	fish swim	this is an example of a dummy sentence that is being used for demonstration purposes	non-entailment	entailment
26	the only animals in the aviary are birds	the aviary does not have a heron living in it	non-entailment	entailment
27	I have never seen a walrus	I have without a doubt seen a walrus	non-entailment	entailment
28	I like Baltimore very much	the moon shone like a burnished medallion	non-entailment	entailment
29	Alice believes Mary is lying	Alice believes Mary	non-entailment	entailment
30	every walrus loves oysters	Paul the walrus loves oysters	entailment	non-entailment
31	my aunt lives in Lagos with her pet walrus	my aunt lives in Lagos	entailment	non-entailment



(O) Infer a Surprise (4/4)

The model's behavior can be explained by a set of generalizations it could have learned from the training examples. On the answer sheet, fill in the blanks to describe these rules. For each rule, also write the numbers of 2 training example sentences (i.e. in the range 1-21) from which it might have learned the rule (there may be more than 2 training examples that could apply):

Rule 1: If __(a)____ is more than __(b)___ words long, label the sentences __(c)_____. (Evidence: example ____(d)____ and example ____(e)_____)

Rule 2: If ___(f)___, label the sentences __(g)_____. (Evidence: example ____(h)____ and example ____(i)_____)

Rule 3: If __(j)____ contains the word _(k)_____, label the sentences __(l)_____. (Evidence: example ____(m)____ and example ____(n)_____)

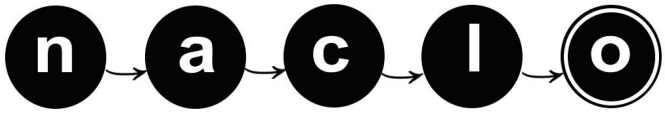
From the examples, it seems that the model has given different priorities to these three rules. Rank the rules in order of priority:

TOP PRIORITY: ____(o)_____

MIDDLE PRIORITY: ____(p)_____

BOTTOM PRIORITY: ____(q)_____

Which test example(s) allow you to determine this ranking? ____(r)_____



(O) Infer a Surprise Answer Sheet

(O) Infer a Surprise

(a) sentence (b) (c)

(d) (e)

(f)

(g) (h) (i)

(j) sentence (k)

(l) (m) (n)

(o) (p) (q)

(r)