1. What's the name of the package you install to get TensorFlow Serving? **1 / 1 point**

- ○ tensorflow-serving
- ○ model-server
- ◉ tensorflow-model-server
- ○ tf-serving

✓ **Correct**

2. What Unix command is used to start TensorFlow serving in a way that will run it and continue running even if the session is disconnected? **1 / 1 point**

- ◉ nohup
- ○ shell
- ○ shellhup
- ○ hup

✓ **Correct**

3. What's the name of the production-scale ML platform for TensorFlow? **1 / 1 point**

- ○ XLA
- ○ TF-E
- ○ TFAX
- ◉ TFX

4. What advantages do you get by running inference on a server instead of distributing the model to all your clients? 1 / 1 point

- Everybody will use the same model version
- You can update your model transparently
- You can scale your model infrastructure for speed and/or volume
- All of the above

✓ **Correct**

5. How do you prepare your model for serving? 1 / 1 point

- Train it and convert it to TensorFlow lite and copy that to the server
- Train it on the server and it will serve automatically
- Give your users the colab URL
- Use TensorFlow SavedModel to save it, and then deploy it to the server

✓ **Correct**

6. If you want to inspect the inputs and outputs for your model, what command do you use? 1 / 1 point

- inspect_saved_model
- Cli_saved_model
- saved_model_cli
- inspect_model

7. If you want to start the model server on port 8501, what parameter do you use? **0 / 1 point**

- (●) --rest_api
- ( ) --rest_api_port
- ( ) --rest_port
- ( ) --api_port

! **Incorrect**

8. I want to pass a list of values (i.e. 8, 9, 10) to the server and have it perform inferences on them, what's the correct syntax for this data? **1 / 1 point**

- ( ) [[8 9 10]]
- ( ) [[8] [9] [10]]
- ( ) [8, 9, 10]
- (●) [[8], [9], [10]]

✓ **Correct**

9. If I publish V1 a model called 'helloworld' and run it with a REST API on port 8501. What's the URL of the endpoint used to run inference on localhost? **1 / 1 point**

- ( ) http://localhost:8501/helloworld:predict/v1
- ( ) http://localhost:8501/models/v1/helloworld:predict
- (●) http://localhost:8501/v1/models/helloworld:predict
- ( ) http://localhost:8501/1//helloworld:predict

10. After running inference using a model hosted on TF Serving, the following is returned. Can you explain what data was sent to the model, and what these return values mean? 1 / 1 point

[[5.77123615e-07, 2.66907847e-08, 4.7217938e-08, 1.97792871e-09, 5.31984341e-08, 0.00734644197, 3.1462946e-07, 0.0439051725, 0.000500570168, 0.948246837],

[0.00227244, 6.12080342e-09, 0.967876315, 3.0579281e-06, 0.0183339939, 3.18483538e-11, 0.011510049, 1.38639566e-14, 4.19033222e-06, 4.40264526e-11],

[1.45221502e-05, 0.999841571, 3.96758715e-08, 0.000131023204, 1.22008023e-05, 1.18227668e-08, 5.97860179e-08, 1.31281848e-08, 5.49047854e-07, 2.97885189e-10]]

- ○ You passed 30 items to the model, and it returned the likelihood that it matches what you need
- ○ Not possible to tell
- ○ You passed an item to a model that recognizes 30 classes and it returned the probabilities for each item that it recognized
- ◉ You passed three items to a model that recognizes 10 classes, and it returned the probabilities for each item in each class

✓ **Correct**