A3.2 - Performance Assessment: theor

Masquer l'énoncé

This task will be graded after the deadline

Votre réponse a passé les tests ! Votre note est de 0.0%. [Soumission #6425a1e0036077b641a41737] ×

Your answers are correctly formatted and have been saved for future grading. You will receive your grade after the deadline.

Question 1: Comparing two models

✓ Your answer is correctly formatted and has been saved for future grading.

A random forest classifies correctly 130 out of 150 test examples.

Compute the 95 % confidence interval for the accuracy of this model.

When rounding, give at least 3 decimals. Example of answer: `0.707, 0.867`.

```
0.812, 0.921
```

Question 2: Comparing two models (continued)

✓ Your answer is correctly formatted and has been saved for future grading.

An SVM with a RBF kernel classifies correctly 115 out of 150 independent test examples.

Compute the 95 % confidence interval for the accuracy of this model.

When rounding, give at least 3 decimals. Example of answer: `0.707, 0.867`.

```
0.699, 0.834
```

Question 3: Comparing two models (continued)

✓ Your answer has been saved for future grading.

Given your answers to the previous questions, can you conclude that the RF model is better than the SVM model?

Beware: you will only receive credit for this question if you answered the two previous ones correctly.

○ Yes

● No

Question 4: Comparing two models (continued)

✓ Your answer is correctly formatted and has been saved for future grading.

We will use a statistical test to decide whether the performances of the two models differ.

Since the sample sizes are here relatively small, we can use a χ^2 test to compare the proportions (rem.: a Fisher exact test is a common alternative likely offering similar numerical results).

What is the p-value of the χ^2 test on these proportions?

When rounding, give at least 3 decimals.

0.037

Question 5: Comparing two models (continued)

✓ Your answer has been saved for future grading.

Can you conclude from this test that there is a statistically significant difference between our two models?

Beware: you will only receive credit for this question if you answered the previous one correctly.

● Yes

○ No

Question 6: Comparing two models (continued)

✓ Your answer is correctly formatted and has been saved for future grading.

What is the minimal number of test examples (instead of the 2 times 150 examples originally used) that would be needed in the previous test to get a p-value below 1%, assuming the test classification rates do not change (86.667 % versus 76.667 %)?

Please report the number of examples needed for each model (not twice this number).

219

Question 7: An evaluation protocol

✓ Your answer has been saved for future grading.

A well informed data analyst observes that a machine learning package is apparently bugged as it produces models that, once tested on independent test examples, have classification accuracies distributed uniformly in the interval $[0\%, 100\%]$.

To support his hypothesis, the data analyst implements the following protocol. He repetitively learns models with the package under study and he observes the accuracies on independent test samples.

More specifically, he reproduces this experiment (learn and test) over several independent training/test sets and he reports as quality measure the *average* of the test accuracies observed on k such test sets. As his results could depend on particular tests, he repeats the whole protocol over 100 distinct runs. He then plots the distribution of this quality measure over the 100 runs and he monitors how this distribution evolves as a function of the number $k = 2, 5, 10, 20, 50, 100$ of test sets considered.

From what you know about the problem at hand, how do you expect the distribution of the quality measure to behave as a function of k?

Select all valid answers.

☑ The quality measure is $\bar{X} = \frac{1}{k}\sum_{i=1}^{k} X_i$.

☐ The quality measure $\bar{X}$ is approximately distributed according $\sim \text{Normal}\left(\mu, \sigma^2\right)$, with $\mu = 0.5$ and $\sigma^2 = \frac{1}{k}$.

☐ The quality measure $\bar{X}$ is approximately distributed according $\sim \text{Uniform}\left(\mu, \sigma^2\right)$, with $\mu = 0.5$ and $\sigma^2 = \frac{1}{k}$.

☑ The quality measure $\bar{X}$ is approximately distributed according $\sim \text{Normal}\left(\mu, \frac{\sigma^2}{k}\right)$, with $\mu = 0.5$ and $\sigma^2 = \frac{1}{12}$.

☐ X_i, the test set accuracy computed on the i th set, is assumed to be distributed according $\sim \text{Normal}(\mu = 0, \sigma^2 = 1)$.

☑ X_i, the test set accuracy computed on the i th set, is assumed to be distributed according $\sim \text{Uniform}(min = 0, max = 1)$.

☐ The quality measure is $\bar{X} = \frac{1}{k}\sum_{i=1}^{k} X_i^2$.

Question 8: An evaluation protocol (continued)

✓

Your answer has been saved for future grading.

Simulate numerically the protocol of the data analyst and check whether these simulations satisfy the expected results from your theoretical analysis of the problem. Some plots should be helpful!

Select all valid answers.

☑ The variance decreases when the number of test sets increases.

☐ The variance remains roughly the same when the number of test sets increases.

☐ The quality measure is not well centered around the expected μ value, found in the previous question.

☐ The variance increases when the number of test sets increases.

☑ The quality measure is well centered around the expected μ value, found in the previous question.

Question 9: An evaluation protocol (continued)

✓

Your answer has been saved for future grading.

What would differ in the previous analysis if the classification test accuracies would be distributed *non-uniformly*, while keeping the same mean and variance?

Select all valid answers.

☐ The variance of the quality measure $\bar{X}$ would stay the same.

☐ The quality measure $\bar{X}$ would no longer be approximately normally distributed. The results are thus expected to be significantly different.

☑ The expected value of the quality measure $\bar{X}$ would stay the same.

☑ The quality measure $\bar{X}$ would still be approximately normally distributed. The results are thus expected to be similar.

Question 10: An evaluation protocol (continued)

✓ Your answer has been saved for future grading.

Based on your answer to the previous question, can you conclude that such a protocol is adequate to assess how the individual test accuracies are distributed?

Beware: you will only receive credit for this question if you answered the previous one correctly.

● No

○ Yes

Soumettre