Computational Statistics

This is a preview of the published version of the quiz

Started: Oct 17 at 4:40am

Quiz Instructions

Aids:

- Non-programmable calculator

Hint:

- Use period as decimal separator, not comma

If you encounter any problems with tables, graphs etc., please use the **PDF version (https://frankfurtschool.instructure.com/courses/11978/files/770792?wrap=1)** **(https://frankfurtschool.instructure.com/courses/11978/files/770792/download?download_frd=1)**

Question 1

1 pts

You are given a two-dimensional Numpy array A:

```
array([[1, 2],
       [3, 4],
       [5, 6]])
```

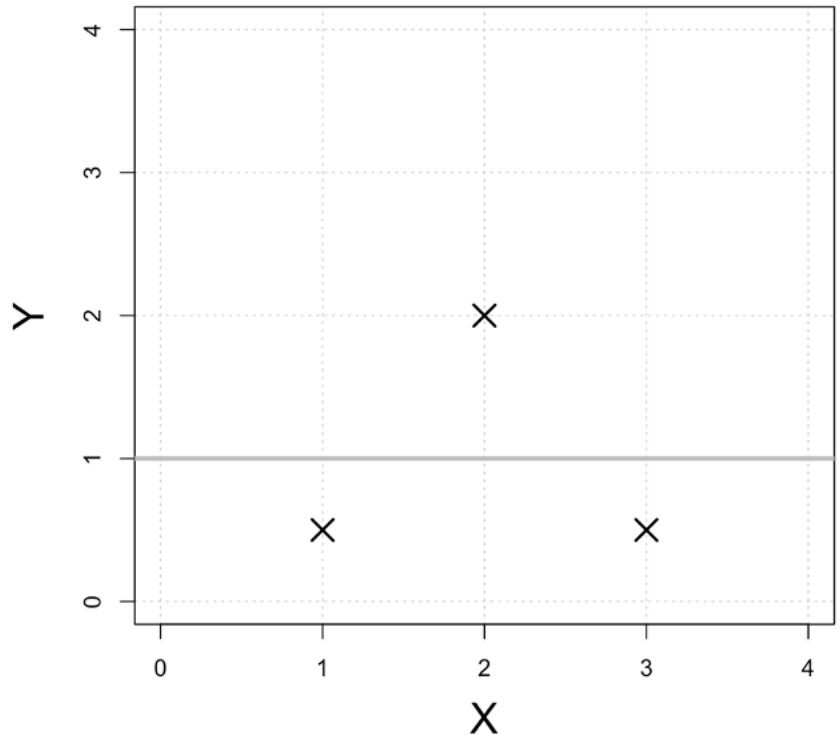
You want variable c to contain the mean of each column, i.e.:

```
array([3., 4.])
```

What needs to be put in for "[...]" in the following code to achieve that?

```
c = np.mean(A, axis=[...])
```

Data frame *df* contains 3 points: $(x_1, y_1) = (1, 0.5)$, $(x_2, y_2) = (2, 2)$, and $(x_3, y_3) = (3, 0.5)$ as shown here:



Question 2

2 pts

What is the SSE (sum of squared errors) of the model shown as the gray line at $\hat{y} = 1.0$ for the entire data set?

Question 3

3 pts

Consider *df* again. Compute the Leave-One-Out Cross Validation SSE for the linear regression model predicting Y from X. Provide your calculations for partial credit.

Edit View Insert Format Tools Table

12pt Paragraph

p

0 words

You collect a set of data points (n = 1000 observations) containing a single feature X and a quantitative target Y. You first split the data into 50% training data *train* and 50% validation data *validation*. Using the training data and Python, you then fit two regression models: a **quadratic regression model** using the single, non-transformed feature X as well as the additional, transformed feature X^2, and a **cubic regression model** using the original feature X as well as additional, transformed features X^2 and X^3.

Question 4

1.5 pts

Suppose that the **true relationship between X and Y is quadratic**, i.e., $Y = b_2X^2 + b_1X + b_0 + \epsilon$ for some b_0, b_1, b_2 and with some Gaussian noise ϵ with $\mathbf{E}[\epsilon] = \mathbf{0}$. Consider the **training MSE** for the quadratic regression and the training MSE for the cubic regression. Would we expect:

- ◯ The error of the cubic regression to be lower.
- ◯ There is not enough information to tell which regression model has lower error.
- ◯ The errors of both regression models to be about the same.
- ◯ The error of the quadratic regression to be lower.

Question 5

1.5 pts

Suppose that the true relationship between X and Y is still quadratic but consider now the **validation MSE** for the quadratic regression and the cubic regression. Would we expect:

- ◯ The error of the cubic regression to be lower.
- ◯ The error of the quadratic regression to be lower.
- ◯ The errors of both regression models to be about the same.
- ◯ There is not enough information to tell which regression model should have lower error.

Question 6

1.5 pts

Suppose now that the **true relationship between X and Y is not quadratic, but we don't know how far it is from quadratic**. Consider the **training MSE** for the quadratic regression and the cubic regression. Would we expect:

- ◯ There is not enough information to tell which regression model should have lower error.
- ◯ The errors of both regression models should be about the same.
- ◯ The error of the quadratic regression to be lower.
- ◯ The error of the cubic regression to be lower.

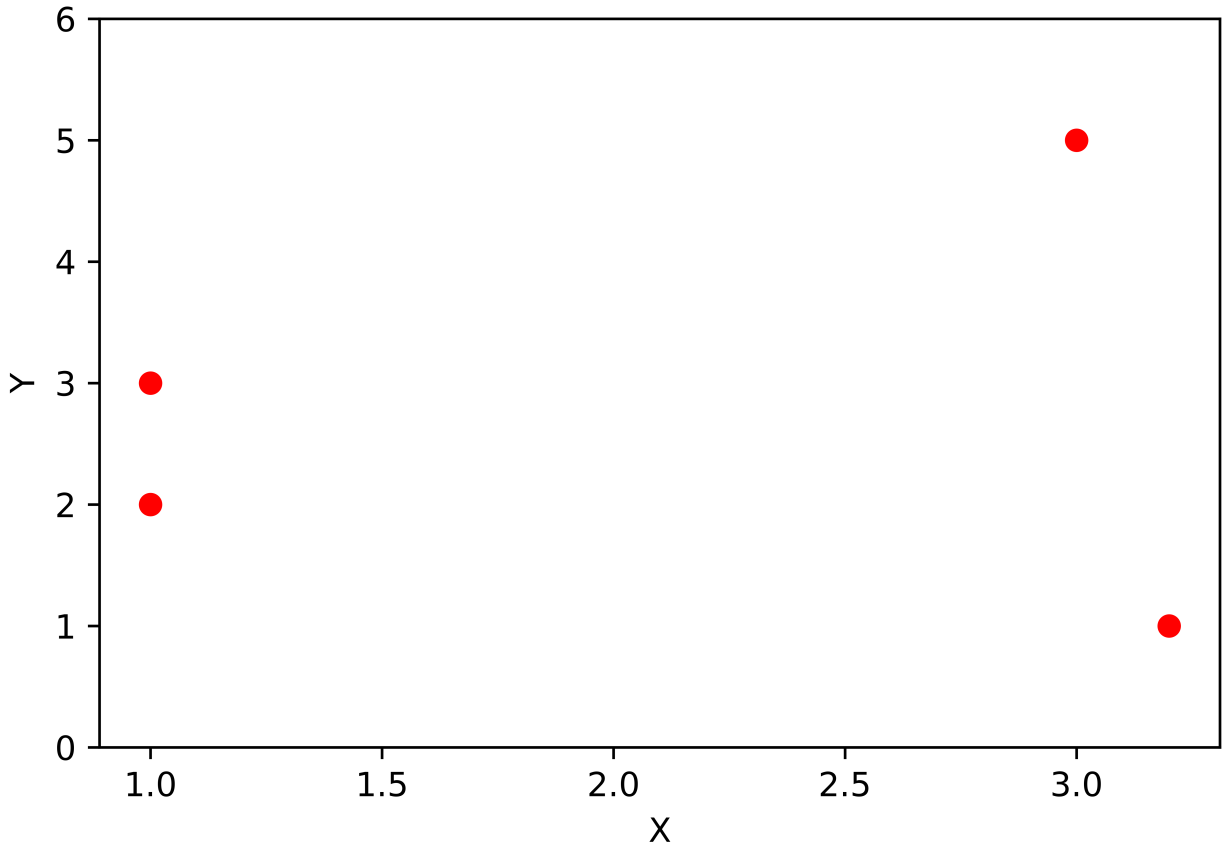
Question 7

1.5 pts

Suppose the true relationship between X and Y is still not quadratic and we still don't know how far it is from quadratic. Consider the **validation MSE** for the quadratic regression and the cubic regression. Would we expect:

- ○ The error of the quadratic regression to be lower.
- ○ The error of the cubic regression to be lower.
- ○ The errors of both regression models should be about the same.
- ○ There is not enough information to tell which regression model should have lower error.

When growing decision trees, the greedy recursive binary splitting algorithm needs to find the optimal split with respect to a given loss function. Consider the following example of a regression task with a single feature X and a target Y (data points are (1,2), (1,3), (3,5), (3.2,1)):



Question 8 — 3 pts

What is the SSE (sum of squared errors) on this training data for the split at X=2.0?

Question 9 — 3 pts

What is the SSE on this training data for the split at X=3.1?

Question 10 2 pts

What is the SSE when splitting at both X=1.5 and X=3.1 on this training data?

Multiple Choice Questions:

Question 11 1 pts

If n is the number of data points, then n-fold cross validation is the same as leave-one-out cross validation.

- ○ True
- ○ False

Question 12 1 pts

You have a dataset with k=3 classes but your classifier can only handle two classes. Extending the binary classifier to k=3 classes using one-vs-all or using one-vs-one requires the same number of binary classifiers to be trained.

- ○ True
- ○ False

Question 13 1 pts

Each tree in a random forest is using a different dataset that was sampled from the original dataset (without replacement).

- ○ True
- ○ False

Question 14 1 pts

The scoring rule $R(y, x) = 2$ is proper.

○ True

○ False

Question 15

1 pts

Even if each *split* in every tree of a random forest only considers a subset of features, it could still happen that every tree is splitting on every feature.

○ True

○ False

Question 16

6 pts

You have trained a bagged ensemble model for regression with B=2 base models on the following dataset:

$(x_1, y_1) = (1, 1), (x_2, y_2) = (2, 4), (x_3, y_3) = (3, 5)$

Your bootstrapped datasets contain the following points:

$D_1 = \{(x_1, y_1), (x_1, y_1), (x_1, y_1)\}$

$D_2 = \{(x_2, y_2), (x_2, y_2), (x_2, y_2)\}$

Let $h_1(x) = 1$ and $h_2(x) = 4$ be the predictions of the first and second base model, respectively.

What is the out-of-bag (OOB) MSE? Give your calculations. (You can write h_1(x_2) or h1(x2) for $h_1(x_2)$.)

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

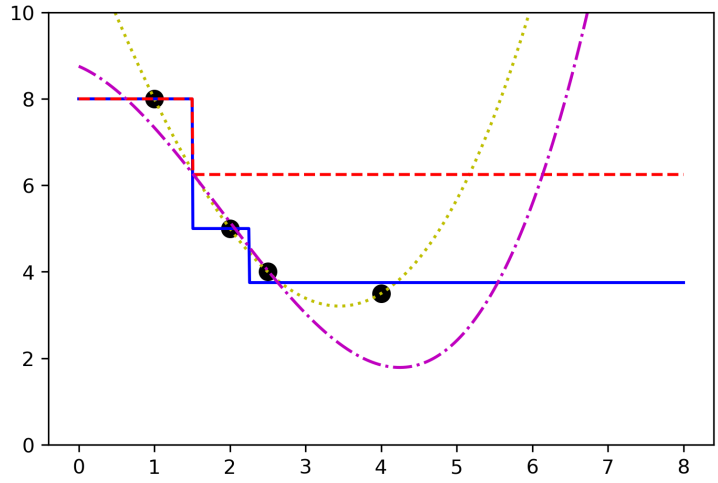
p

0 words </>

Question 17

6 pts

This is a plot of the four training data points and the predictions of four different regression models that were trained on (only) these points. There is a **1-to-1 mapping** from the prediction lines to the models. Which line belongs to which model?



Bagged decision tree model	[Choose]
Decision tree model	[Choose]
Polynomial of degree 3	[Choose]
Bagged model of polynomials (degree 3)	[Choose]

You are a probabilistic forecaster competing with two other forecasters for a single prize that will be given to the forecaster with highest logarithmic scoring rule score. (Ties will be broken uniformly at random, so that all forecasters with highest logarithmic score receive the prize with equal probability.)

The logarithmic scoring rule is: $R_l\left(y,\ x\right)\ =\ x\ln(y)+(1-x)\ln(1-y)$ (which becomes $\ln(y)\ \text{if}\ x=1$ and $\ln(1-y)\ \text{if}\ x=0$).

The forecasting competition has only a single question and you believe that the event will happen with probability p = 0.3. You know that the other two forecasters also believe that the probability of the event occurring is 0.3. Moreover, you know that they will report truthfully.

Question 18

1 pts

What is your subjective **probability of being selected** if you report 0.6?

- ○ 1
- ○ 0.6
- ○ 1/2
- ○ None of these
- ○ 0.3
- ○ 0.7

Question 19

2 pts

Which of these **forecasts maximizes your subjective probability of winning the prize?**

- ○ 0.1
- ○ None of these
- ○ 0
- ○ Any of these

Question 20

2 pts

Which of these forecasts maximizes your **subjective probability of winning the prize if the other two forecasters are not truthful but report 1.0 (i.e., 100%) instead?**

- ○ 0.0
- ○ Any of these
- ○ 0.5
- ○ None of these

Question 21

2 pts

What is the **highest expected probability of winning the prize that you can obtain assuming the other two forecasters report 1.0 (i.e., 100%)?**

Question 22

2 pts

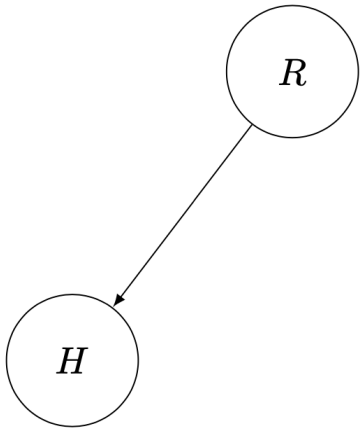
In the lecture, we have seen that we can implement truthful forecasting competitions by giving the prize to forecaster i with probability

$$f_i = \frac{1}{n} + \frac{1}{n}\left(R_q(y_i, x) - \frac{1}{n-1}\sum_{j \neq i} R_q(y_j, x)\right),$$

where $R_q(y, x) = 1 - (y - x)^2$.

For $n = 5$, what is the highest possible selection probability for forecaster 1, i.e., what is $\max_{x, y_1, y_2, y_3, y_4, y_5} f_1$?

Consider the following Bayesian network representing the beliefs of a weather forecaster seeking to predict the probability of rain (R). The forecaster has been offered access to a (noisy) hygrometer (H) with the conditional probability table shown in the graphic.



$$\Pr(R=1)=0.7 \quad \Pr(R=0)=0.3$$

	H	
R	1	0
1	0.8	0.2
0	0.6	0.4

Question 23 2 pts

What is the forecaster's belief that the hygrometer signal will be 1? That is, what is Pr(H=1)? (use **at least three digits after the decimal point** in your calculations wherever possible)

Question 24 2 pts

What is the forecaster's belief of rain given that the hygrometer signal is 0? That is, what is Pr(R=1|H=0)? (use **at least three digits after the decimal point** in your calculations wherever possible)

Question 25 4 pts

What is the expected quadratic score of the truthfully-reporting forecaster if she expects to learn the outcome of H (i.e., she will have access to the hygrometer but does not know its value yet)? (The numerical result alone is sufficient but you can provide your calculation for partial credit; use **at least three digits after the decimal point** in your calculations wherever possible.)

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

Question 26

6 pts

You are supposed to predict the probability that a fair die comes up as a "6" and you are scored by the logarithmic scoring rule.

Assume someone could tell you whether the number that came up was "less than 5" (1,2,3,4) or "at least 5" (5, 6) before you need to report a probability. How much would this piece of information be worth to you?

Let $0 \cdot \ln(0) = 0$ and provide your calculations using **at least three decimals** wherever possible.

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

Saving... Submit Quiz