Time left 1:47:05

Question **1**

Not yet answered

Marked out of 12.00

Suppose that you have the following results of survey. Now it is necessary to analyze obtained survey by using Simple Random Sampling and Clustered Random Sampling. Recommended tool for using is MS Excel.

(Ignore the fpc and the clustering in calculating the standard error.)

P.S. when you are going to write your answer into answer sheet, please round up to 2 digits after floating point

1) Assume that you're going to do Simple Random Sampling (SRS) for above dataset.

Compute a mean: ☐☐☐☐

P.S here you need to drag-and-drop digit-by-digit. For example, if your answer is 35.13 then you need to drag 3, then 5, then . and then 1 and 3

2) Compute a standard error for SRS: ☐☐☐☐

3) Now compute 95% of confidence interval. Please note that t-value in this case is equal to 2.04

Upper limit for SRS: ☐☐☐☐☐

Lower limit for SRS: ☐☐☐☐☐

P.S You need to take **rounded** answers from 1st and 2nd questions

Assume that you're going to do Clustering Random Sampling for above dataset.

1) Compute a mean: ☐☐☐☐

2) Compute a standard error for Clusteting Random Sampling: ☐☐☐☐

3) Compute d-value: ☐☐☐☐

Hint: d- value is a ratio of standard error for clustering over standard error for SRS

4) Compute d-squared: ☐☐☐☐

5) Compute roh: ☐☐☐☐

P.S In 5-th step: You need to take answers from 3rd and 4th questions

Hint: W_{cl} =0.125 And you need to calculate (S_{cl1}...S_{cl8}) 8 times for each cluster.

6) N_{eff} = ☐☐☐☐

1 2 3 4 5 6 7 8 9 0 ,

Question **2**

Not yet answered

Marked out of 10.00

Suppose that you're going to run linear regression with some input features and 1 output feature. Your hypothesis is

$$h_\theta(X) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_1^2 + \theta_4 X_1^3 + \theta_5 X_2^2 + \theta_6 X_2^3 + \theta_7 (X_1 \cdot X_2) + \theta_8 (X_1^2 \cdot X_2)$$

```
X = np.c_[np.ones(df.shape[0]), df[['X1', 'X2', 'X1^2', 'X1^3', 'X2^2', 'X2^3', 'X1*X2', 'X1^2*X2']].values]
Y = df['Y'].values.reshape(-1, 1)
```

Firstly it it necessary to normalize your dataset: Z = (x-mu)/std

Initial theta parameters is equal to zero. Learning rate is 0.1. Now, let's complete the following table:

#Iterations	Cost Function (Round please up to integer value)	Optimal Theta parameter Indicate here maximum theta value(Round please up to integer value)
n=10		
n=100		
n=1000		

1 2 3 4 5 6 7 8 9 0

Question **3**

Not yet answered

Marked out of 10.00

Suppose that you have the following dataset, with 3 input features, and 1 output feature. You're going to apply Logistic Regression algorithm with regularization.

Firstly it is necessary to apply normaization with the following formula: Z = (X-mu)/std.

Initial theta parameters = 0.

#Iterations, lambda, learning rate	Cost function (rounded up to 2 digis after floating point)	Optimal theta parameter Indicate here maximum theta value (rounded up to 2 digis after floating point)
N=100, alpha = 0.1, lambda = 0.1		
N=1000, alpha = 0.2, lambda = 1		
N=10000, alpha = 0.3, lambda = 10		

After 10.000 iterations, alpha = 0.3, lambda = 10 and by setting threshold = 0.5, what is the number of ones in the first 10 rows of prediction:

1 2 3 4 5 6 7 8 9 0 ,

Question **4**

Not yet answered

Marked out of 10.00

Suppose that you're going to run neural network algorithm with the following parameters:

```
1. Data Preparation: Normalized input vectors for binary classification.
2. Network Architecture: 3 hidden layers with Tanh, output layer with Sigmoid.
3. Forward Propagation: Computes activations through layers.
4. Loss Calculation: Uses Mean Absolute Error (MAE) as the loss function.
5. Backpropagation: Computes gradients using chain rule and Tanh derivative.
6. Weight and Bias Updates: Uses gradient descent with learning rate 0.1.
7. Iterative Training: Runs for 10,000 epochs, prints loss every 1000 epochs.
8. Final Prediction: Generates probabilities for each input.
```

a4 = [□□□, □□□]

a3.min() = □□□□ (round up to 3 digits after floating point)

W4.max() = □□□□ (round up to 2 digits after floating point)

W3.min() = □□□□ (round up to 2 digits after floating point)

Loss after 10000 epochs: □□□

General Conculsion after 10000 epochs:

1 2 3 4 5 6 7 8 9 0 ,

NN predicts image of dog | NN predicts image of cat | NN can't define correct image class

Question **5**

Not yet answered

Marked out of 9.00

For a multi-class classification problem, we don't calculate an overall F-1 score. Instead, we calculate the F-1 score per class in a one-vs-rest manner. In this approach, we rate each class's success separately, as if there are distinct classifiers for each class.

Suppose that you are running logistic regression with some threshold and now you have the following in Table 1.

Table 1

		Actual Class		
		a	b	c
Predicted Class	a	30	20	10
	b	50	60	10
	c	20	20	80

a) Firstly, calculate the accuracy of algorithm

b) Find precision and recall for each classes a, b and c. Hint:

c) Calculate F1-score for each class a, b and c.

For part b and c please transfer your answers into table below:

Class	Precision	Recall	F1-score
a			
b			
c			

Please round up to 3 digits after floating point

1 2 3 4 5 6 7 8 9 0 ,

◄ Supplementary Materials (7-variant)

Jump to...