Quiz Questions

1 Simple Linear Regression

Question 1. How did we arrive at the equations for the slope and intercept for the linear regression model?

Question 2. How does the correlation coefficient factor into your understanding of linear regression?

Question 3. How do you fit an exponential curve using the linear regression model? How do you transform the data to get a good fit?

2 Multiple Linear Regression

2.1 Ordinary Least Squares

Question 1. What types of Date need the multiple linear regression analysis?

Question 2. In a multiple regression when two independent variables are correlated, this is referred to as _____and when three or more variables are correlated, this is referred to as _____.

A. regression, correlation;

B. collinearity, multicollinearity;

C. correlation, regression;

D. multicollinearity, collinearity;

Question 3. In multiple regression with p predictor variables, when constructing a confidence interval for any β_i, the degrees of freedom for the tabulated value of t should be:

A. $n-1$

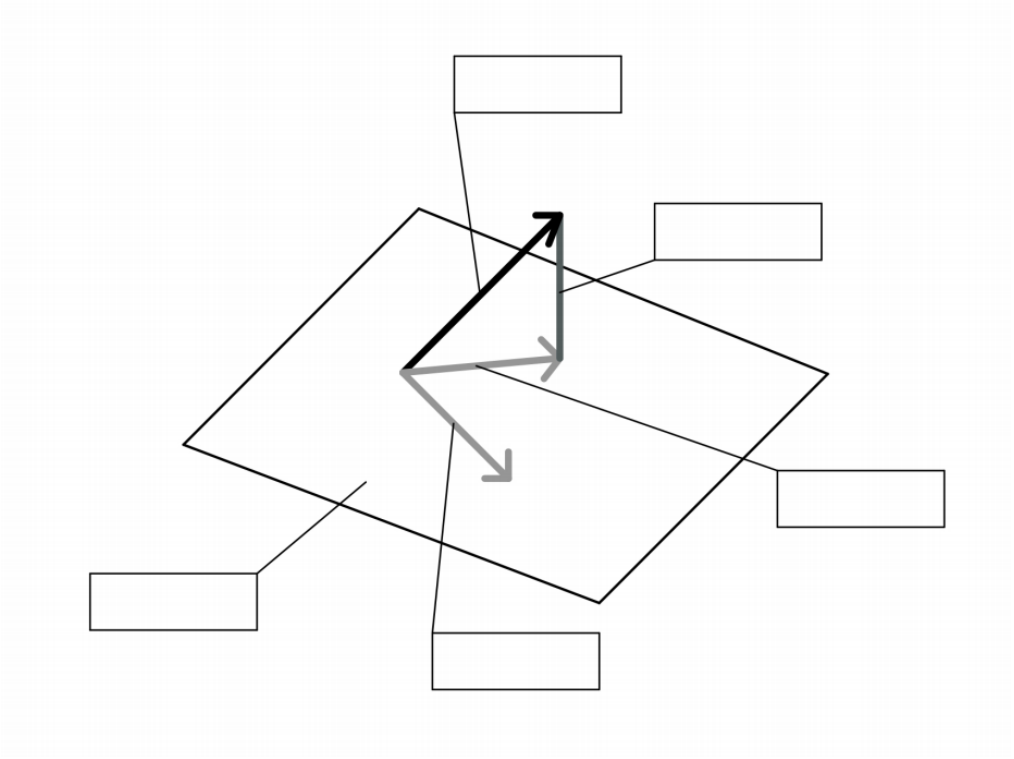
B. $n-p-2$

C. $n-p-1$

D. $n-p$

2.2 Geometric Perspective

Suppose we have a dataset represented with the design matrix span($\mathbb{X}$) and response vector $\overrightarrow{y}$. We use linear regression to solve for this and obtain optimal weights as $\hat{\beta}$. Draw the geometric interpretation of the column space of the design matrix span($\mathbb{X}$), the response vector $\overrightarrow{y}$, the residuals $\overrightarrow{y}$ - $\mathbb{X}\hat{\beta}$, and the predictions $\mathbb{X}\hat{\beta}$.



(a). What is always true about the residuals in least squares regression? Select all that apply.

☐ A. They are orthogonal to the column space of the design matrix.

☐ B. They represent the errors of the predictions.

☐ C. Their sum is equal to the mean squared error.

☐ D. Their sum is equal to zero.

☐ E. None of the above.

3 Categorical Data and One-Hot Encoding

Question 1 What is categorical data? Give an applicable example of when encoding categorical data would be useful and why.

Question 2 Draw an example of a table before and after one-hot encoding.

Question 3 Fill in the code blanks.

```
from sklearn._________ import LinearRegression
from sklearn._________ import OneHotEncoder

dat = pd.DataFrame({"color": [red, green, blue, red, pink, black, yellow],
                    "article": [shirt, shoes, jacket, jacket, pants, pants, shirt]}

# create encoder

enc = _______________________________

# fit encoder

___________________________________

# classify 'x'
x = [['red', 'shirt'], ['pink', 'jacket'], ['rainbow', 'horse']]

enc._________________(x)
```

Question 4 From the previous question, what does the vector result of x look like?

```
    [[__,__,__,__,__,__,__,__,__,__],
    [__,__,__,__,__,__,__,__,__,__],
    [__,__,__,__,__,__,__,__,__,__]]
```

4 Features

Question 1 What feature map should be created to model the equation $y = 5x^2 + 1$?

Question 2 True or False. We can use least squares to perform regression and approximately find a such that y = sin(ax).

5 Linear Regression for Classification

5.1 Binary Classification

Question 1

We have learned a weight vector $\mathbf{w} = [3.1, 0.7]^T$. What class (1 or -1) would we assign to the data point $\mathbf{x} = [-0.5, 1]^T$.

Question 2

True or False. If some binary class data is linearly separable, linear regression will always find a separating hyperplane.

5.2 Multivariate Classification

Question 3

Suppose we had a labeled dataset of 10,000 sentences written in different languages. We wish to classify the language of each sentence as either English, French, German, or Spanish based on the number or occurrences of each letter in the sentence (assume we are using 30 different characters). We are using linear regression with the one-vs-all strategy. How many parameters would we need to learn from the data?