Quiz Solutions

1 Simple Linear Regression

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

Answer: We took the error squared from the equation $y = mx + b$ and calculated the derivative being zero for partial derivatives of m and b.

Thus we arrived at $\sum y_i - m \sum x_i - bn = 0$ and

$\sum y_i x_i - m \sum x_i^2 - b \sum x_i = 0$.

Solving out, we got $b = \sum y_i - m \sum x_i / n$ and

$m = \sum x_i y_i - \sum y_i \sum x_i / n / \sum x_i^2 - \sum (x_i)^2 / n$

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

Answer: The correlation coefficient plays into understanding for Question 1 as well as visually seeing the representation of linear correlation between x and y data.

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?

Answer: You have to transform the y data in the model to get the equation $ln(y) = mx + b$ as appropriate from the original curve $y = e^{mx+b}$. In this way, you can get a good fit for a linear regression model.

2 Multiple Linear Regression

2.1 Ordinary Least Squares

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

Answer: Multiple continuous X variables and continuous y response.

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;

Answer: B

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. $n - p - 2$

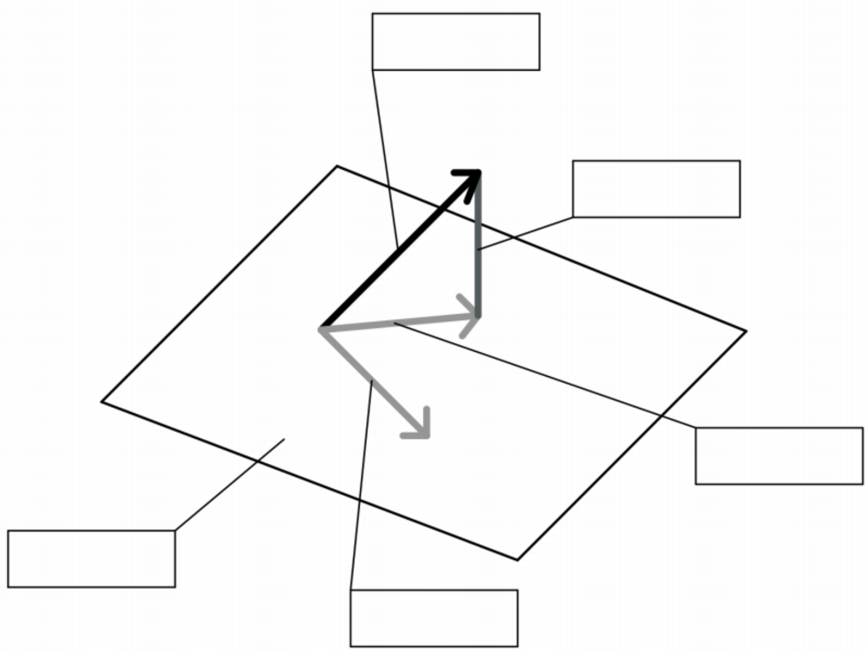
C. $n - p - 1$

D. $n - p$

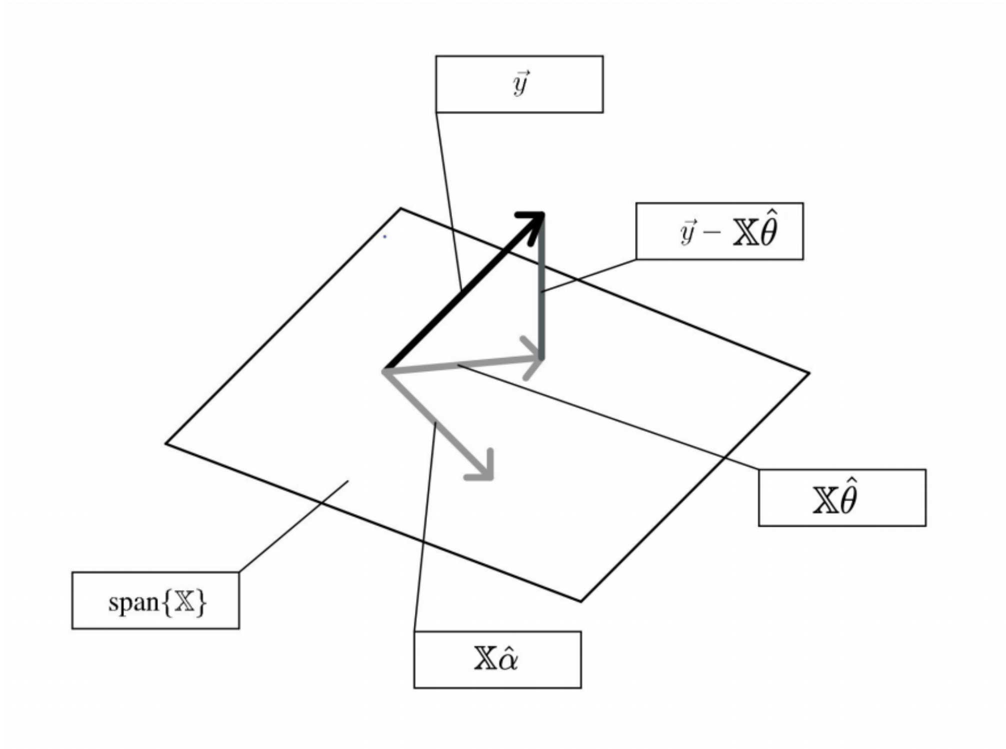
Answer: C

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 $\mathbb{X}\hat{\beta}$.



Answer:



(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.

Answer: A and B

C is wrong because the mean squared error is the mean of the sum of the squares of the residuals. D: A counter-example is: $\mathbb{X} = \begin{bmatrix} 2 & 3 \\ 1 & 5 \\ 2 & 4 \end{bmatrix}$, $\mathbb{Y} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ After solving the least square problem, the sum of the residuals is 0.0247, which is not equal to zero. However, note that this statement is in general true if every feature contains the same constant intercept term. E is wrong since A and B are correct.

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.

Answer: Categorical data is a data type that considers non-numerical, un-ordered data that creates categories. Categorical data can be useful for collecting and using data for mortality rates of countries while considering geographic region, access to healthcare, or gender.

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

Answer should resemble this table:

state
NY
WA
CA

AL	...	CA	...	NY	...	WA	...	WY
0	...	0	...	1	...	0	...	0
0	...	0	...	0	...	1	...	0
0	...	1	...	0	...	0	...	0

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)
```

Solution: blue

```
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import OneHotEncoder

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

# create encoder

enc = OneHotEncoder(handle_unknown='ignore')

# fit encoder

enc.fit(dat)

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

enc.transform(x)
```

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

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

Answer:

```
[[1, 0, 0, 0, 0, 0, 1, 0, 0, 0],
[0, 0, 0, 1, 0, 0, 0, 0, 1, 0],
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]
```

4 Features

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

Answer: We should create a polynomial feature map. Specifically, we can create a univariate degree 2 polynomial feature map. I.e. $\phi : x \to (1, x, x^2)$. If we wanted to make the modelling process even easier, we can force the coefficient of x to be zero by creating the feature map $\phi : x \to (1, x^2)$.

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

Answer: True. We know that the sin function can be taylor expanded in terms of polynomial features, so we can do univariate polynomial regression with a high degree and then reverse engineer a.

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 $\mathbf{x} = [-0.5, 1]^T$.

Answer:. $\mathbf{w}^T\mathbf{x} = -0.85$, so we would assign this point to the -1 class.

Question 2

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

Answer: False. In the notebook, we saw that if we have an imbalanced class the decision boundary can be skewed such that points are misclassified even though the data is linearly separable. This is because correctly classifying all points does not mean the MSE is minimized.

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?

Answer: We need to fit a model that separates each class from the rest. Since we have 4 classes, we would have 4 weight vectors of 31 coefficients (30 features + 1 bias term) each for a total of 124 weight coefficients.