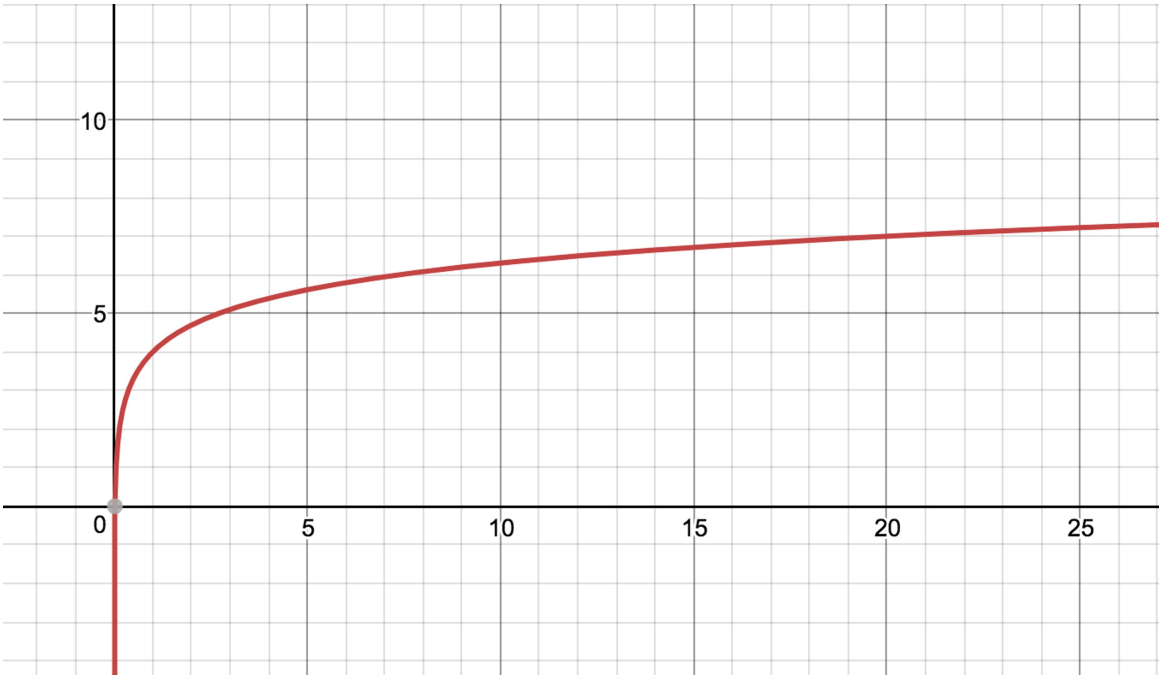
Q&A

05 February 2020 19:43

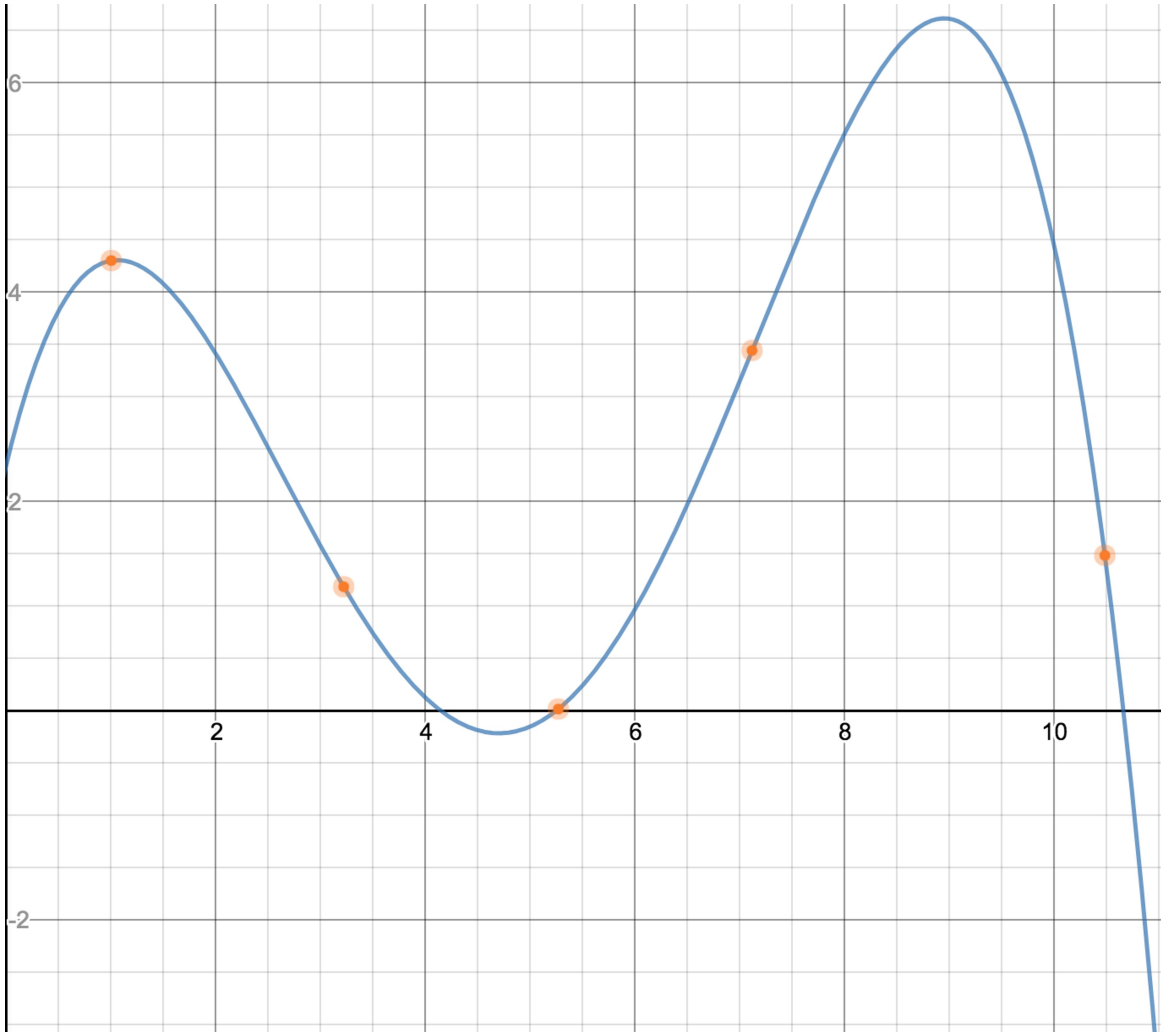
1.) Identify the graph
Below is the graph (x-axis: months, y-axis: users (in millions)) is the rate at which consumers are signing up for the "YouFood" on a monthly basis. Initially, people were signing up for this service at a very rapid rate, as within two months it had a consumer base of 4.8 million users. But after fewer months with the arrival of competitors i.e. other food delivery startups, the consumer base is still increasing but with a significantly lower rate. (e.g. in 10 months it had ~6.3 million users and in 20 months it has increased to ~7 million only)



Which of the following functions approximately describes the above curve?
Y=4+ln(x)
Feedback :
As natural logarithm function slowly grows to positive infinity as x increases but for the sudden increase there should be some positive constant hence 4+ln(x) describes the above phenomena

2.) Sales Figure
The sales graph (discussed in the lecture) shows a wavy nature. Select the appropriate function which could possibly depict wavy nature.
Y ~ C cos(x)
Feedback :*The graph of cosine/sine function repeats itself at intervals of 360 degrees with one peak and one dip (in each 360 degree interval) giving it a wavy nature.*

3.)
Profit Function
Which of the following equations best represents the following graph?



ax4+bx3+cx2+dx+f
Feedback :
Functions having no global maxima or minima are usually polynomial functions. Also, they typically have multiple roots and local maxima and minima.

4.) Feature Generation
Can x1.x2.x3 be a feature if the raw attributes are x1, x2, x3 and x4?
Yes
Feedback :
Derived features can be created using any combination of the raw attributes (linear or non-linear). In this case, the combination x1. x2. x3 is non-linear.

5.) Feature Generation
How many maximum features can be created if we have d raw attributes for n data points? Note that (nr) here refers to the number of ways of selecting r items from a set of n.
Infinite
Feedback :
You can (in principle) create as many features as you want.

6.) Consider x1 and x2 as two independent variables which are used to create certain features. Which of the following is a non linear feature? Select the correct option.

X1 *X2
X1/X2
Both B and C
As order is not 1, linear feature has order 1, so X1+X2, X1-X2 is linear.

7.) Raw Attributes vs Features
Look at the following table:

Cost of House	Date of Establishment	Number of Rooms	Carpet Area	Area per room(Carpet Area/Number of Rooms)
120	10/2/2004	3	1410	470
100	23/1/1998	2	950	475
195	1/13/2001	5	2100	420

We are trying to predict the Cost of House on a test set. The sample of the training set is shown in the above table.
Which of the above variables do you think are raw attributes?
The units of some of the attributes are as follows:
Cost of House: lakhs of INR
Date of Establishment: dd/mm/yyyy
Carpet Area: sq.ft

Date of Establishment
Feedback :
It is not derived
Number of Rooms
Feedback :
It is not derived
Correct
Carpet Area
Feedback :
It is not derived

8.) Raw Attributes vs Features
Look at the following table:

Cost of House	Date of Establishment	Number of Rooms	Carpet Area	Area per room(Carpet Area/Number of Rooms)
120	10/2/2004	3	1410	470
100	23/1/1998	2	950	475
195	1/13/2001	5	2100	420

We are trying to predict the Cost of House on a test set. The sample of the training set is shown in the above table.
Which of the following options can be used as features?
Note:
Cost of House: lakhs of INR
Date of Establishment: dd/mm/yyyy
Carpet Area; sq.ft
Date of Establishment
Feedback :
Raw Attributes, as well as the Derived Features, can be used as Features
Correct
Number of Rooms
Feedback :
Raw Attributes, as well as the Derived Features, can be used as Features
Correct
Carpet Area
Feedback :
Raw Attributes, as well as the Derived Features, can be used as Features
Correct
Area per room
Feedback :
Raw Attributes, as well as the Derived Features, can be used as Features

9.) Feature Generation
Raw Attributes vs Features
Look at the following table:

Cost of House	Date of Establishment	Number of Rooms	Carpet Area	Area per room(Carpet Area/Number of Rooms)
120	10/2/2004	3	1410	470
100	23/1/1998	2	950	475
195	1/13/2001	5	2100	420

We are trying to predict the Cost of House on a test set. The sample of the training set is shown in the above table.
How can we use Date of Establishment as a feature?
Note:
Cost of House: lakhs of INR
Date of Establishment: dd/mm/yyyy
Carpet Area; sq.ft
Suggested Answer
We can subtract Date of Establishment from the current date to create the age of the house as a feature.

10.) Vector Product

Consider two vectors a = (a1,a2) and b= (b1,b2). Find the dot product of the vectors - (a*b). Select the correct option.
a1b1 + a2b2

11.)Linearity
Is the following equation linear?
y=ax1+bex2+x3+csin(dx4)
Note : a,b,c and d are coefficients of regression.
False

12.)Linearity
If no, why is it not linear?
y=ax1+bex2+x3+csin(dx4)
Note : a,b,c and d are coefficients of regression.
Because of the term csin(dx4)
Feedback :
As stated in the text and the video, the explanatory variable should be linear with respect to the coefficients.

13.) Value of Hyperparameter
When we choose the value of the hyperparameter close to zero, it might lead to overfitting.
True
Feedback :
λ = 0 implies that there is no regularization and there is a high chance of overfitting.

14.) Hyperparameter
Which of the following is the hyperparameter in the objective function that we minimize for regularized regression?
λ
Feedback :
It is the hyperparameter.

15.) Computational Advantage

Which of Ridge and Lasso regressions is computationally more intensive?
Lasso Regression
Feedback :
Ridge regression almost always has a matrix representation for the solution while Lasso requires iterations to get to the final solution.

16.) Variable Selection
Which of the following methods perform variable selection, i.e. can help discard redundant variables from the model?
Lasso Regression
Feedback :
Lasso trims down the coefficients of redundant variables to zero, and thus indirectly performs variable selection also. Ridge, on the other hand, reduces the coefficients to arbitrarily low values, though not zero.

17.) Cost Function in Regularised Regression
The cost function in regularized regression models has two terms - the error term and the regularization term. The objective of the learning algorithm is to find the coefficients alpha such that:
The sum of the error term and the regularization term is minimised
Feedback :
The objective function to minimise is the sum of the error and the regularization terms.

18.) Graphical Interpretation of Ridge and Lasso Regression
The red contours (in the above image) depict the graph of the error term. Select the correct order of the values of error for the (red) contours. The innermost one is labelled as 3, then 2, then 1.
1>2>3

19.) Optimal Values of the Coefficients
In the contour plot shown, the axes represent the coefficients of the model $\alpha 1$ and $\alpha 2$. The contours represent how the error and the regularization terms vary with the coefficients.
What kind of a model will you get if both the coefficients have extremely low values, i.e. both the coefficients are close to zero?
A simple model which is likely to underfit the data
Feedback :
When both coefficients are close to zero, you are somewhere near the origin, where the error contours are high but the regularization contours are very low. Such a model is likely to underfit the data since the error is quite high (though it is 'simple' since the coefficients are small).

20.)Feature Selection using VIF and p-value
A regression exercise is performed on yy using the variables x1, x2 and x3. The VIF and the p-value after the first iteration are shown as below:

Variable	VIF	p-value
x1	4.2	0.001
x2	3.1	0.09
x3	1.8	0.04

Which variable would you remove and reiterate?
x2
Feedback :
It has a VIF value greater than 3 and the significance is also greater than 0.05, hence it should be eliminated first.

21.)Comparison of Cp and BIC
Suppose we are regressing using least squares an independent variable y on 20 predictors, which one of BIC and Cp will have a lower value?
It depends
Feedback :

If the number of observations, n > 7, then Cp will give a lower value else BIC.

22.) Adjusted R2
Value
Can Adjusted R2 be negative?

Yes

Feedback :

Adjusted R^2 can be negative if $\frac{RSS/(n-d-1)}{TSS/(n-1)}$ is greater than 1, which implies $\frac{RSS}{n-d-1} > \frac{TSS}{n-1}$, which is possible when the predictors do not explain the dependent variable at all such that RSS ~ TSS.

23.)Best Subset Selection: Count of Models
How much is (10 2)?
45
Feedback :
The number of ways of choosing 2 apples from 10

24.) Best Subset Selection: Count of Models
Suppose there are 14 predictors to build a model. How many models can be built using 0 predictors, 1 predictor, 2 predictors and 3 predictors?
1, 14, 91 and 364 respectively
Feedback :
Look at the previous question and figure out what (10 2) means

25.)

Best Subset Selection: Count of Models

How many total models can be built using 10 predictors?

- ☐ 3^{10}
- ☑ 2^{10} ✓ Correct
  Feedback :
  Try to generalise the above formula as a summation from 1 to 10 for the number of predictors. What is $\sum_{k=0}^{n} \binom{n}{k}$? Think of binomial expansion of $(x+y)^n$.
- ☐ 59049
- ☑ 1024 ✓ Correct
  Feedback :
  Try to generalise the above formula as a summation from 1 to 10 for the number of predictors. What is $\sum_{k=0}^{n} \binom{n}{k}$? Think of binomial expansion of $(x+y)^n$.

26.) Stepwise Selection: Count of Models
In Forward Stepwise Selection, how many models do we fit at each step as we increase d? The total number of predictors is p.
p,p−1,p−2,.....,2,1
Feedback :

Look at the flowchart for Forward Stepwise Selection

27.) Stepwise Selection: Count of Models
In Backward Stepwise Selection, how many models do we fit at each step as we decrease d? The total number of predictors is p.
p,p−1,p−2,.....,2,1
Feedback :
Look at the flowchart for Backward Stepwise Selection

28.) Stepwise Selection: Count of Models
How many models do we need to fit for Forward Stepwise Selection?
$1 + \frac{p(p+1)}{2}$
Feedback :
Good that you considered the null model.

29.) Method Selection
Now, suppose we are regressing using 120 predictors on a dataset of 80 observations. Which of the following methods can be used to perform regression?
Forward Stepwise Selection
Feedback :
When n < p and the number of predictors are > 40, Forward Stepwise selection can only be used.

30.) Best Model Selection
Suppose we are regressing an independent variable y on 18 predictors on a dataset with 400 observations. Which method shall be able to give us the best model, that is with lowest test error?
Best Subset Selection
Feedback :
Best Subset Selection can get the best model as it tries each and every combination and here, the number of predictors is also less than 40.