Q&A

02 February 2020 17:29

1.)Central Issue in Machine Learning
The central issue in machine learning can be said to be the study of:

How to extrapolate learnings from a finite amount of data to explain or predict all possible inputs of the same kind
Feedback :*Machine learning does not simply involve building models to fit the available data. The real challenge is to learn patterns which can be used to explain the behaviour of similar unseen data.*

2.) Occam's Razor
Occam's razor is a fundamental principle which suggests that:
A model should be as simple as possible, but robust
Feedback :*Occam's razor does not say that a model should be unjustly simplified till no further simplification is possible. It says that when faced with a trade-off between a complex and a simple model, with all other things being roughly equal, you are better off choosing the simpler one. The reason for this will be explained shortly.*

3.) Regression Models
Choose the simplest regression model among the following (all lowercase alphabets are features):
Y = x + 3w

4.) Model and Learning Algorithm
The relationship between a model and a learning algorithm is that:
A learning algorithm learns from training data and produces a model
Feedback :*Learning algorithm's task is to figure out what needs to be done and how. In linear regression, for example, the learning algorithm optimises the cost function and produces the models, i.e. the coefficients.*

5.) Linear Class of Models
A learning algorithm belonging to the 'linear' class will only produce linear models. Which of the following models can NOT be produced by such a learning algorithm?
k-NN
Feedback :*Logistic regression is a linear model. Recall that log(odds) is equal to a linear expression in the features. The full form of glm() is generalised linear models. K-NN is non-linear (Food for thought: Is it even a model or just a learning algorithm?)*

6.) Model and Learning Algorithm
The learning algorithm is told what needs to be done, it figures out how it needs to be done and returns a model

7.) Simplicity and Complexity
The reason simpler models are considered better than complex models are (mark all the correct options):
Simpler models are generic i.e. they apply to a wider range of data
Complex models make assumptions about the data which are likely to be wrong
Simpler models require less training data than complex models
Simpler models are more robust

8.) Overfitting
The most extreme case of overfitting is:
The first person has mugged up all the possible questions using numerous text-books and

preparation material

9.) Disadvantages of Complexity
The disadvantages the first person is likely to face because of a complex model are (mark all that apply):
He'll need more training data to 'learn'
Despite the training, he may as well not learn and perform poorly in the real world

10.) Overfitting
The possibility of overfitting exists primarily because:
Models are trained on a set of training data but their efficacy is determined by the ability to perform well on unseen (test) data
Feedback :*It is possible to memorize the training data while failing to truly learn the underlying trends and patterns. On unseen data (read tricky but unseen exam questions), memorizing is bound to fail.*

11.) Overfitting in Linear Regression
In linear regression, which of the following are clear signs of overfitting?
The R-squared value on training data is 0.90 and 0.30 on train and test data respectively
Feedback :*In overfitting, the model fits the training data very well since it has somehow memorised it.*

12.)Identifying Overfitting in a Neural Network
Say you have a 100,000 Google images as training observations and you are trying to build a neural network to classify the images in 3 classes - nature, cities and others. You use another 50,000 observations to test it and the accuracy on the test set comes out to be 10%. Which of the following can you use to check whether the model has overfitted?
Accuracy on the training set
Feedback :*If the training accuracy is way higher than 10%, it is likely to have overfitted. Neural networks, as you'll study later, can be made extremely complex using a number of hyperparameters like hidden layers and number of neurons. The basic idea, though, remains the same - memorisation of training and and failure on test data.*

13.) Bias and Variance
We said that the first person's mental model has high variance and that the second one's has high bias. Relating model complexity to bias and variance, we can say that:
Complex models have high variance and simple ones have high bias
Feedback :*The first person's model is complex, since he/she needs to memorise a lot, and is therefore prone to change when the exam pattern or syllabus (read training data) changes. The second one's model is simple and may be too simple to crack a competitive exam, hence high bias (read 'naive model').*

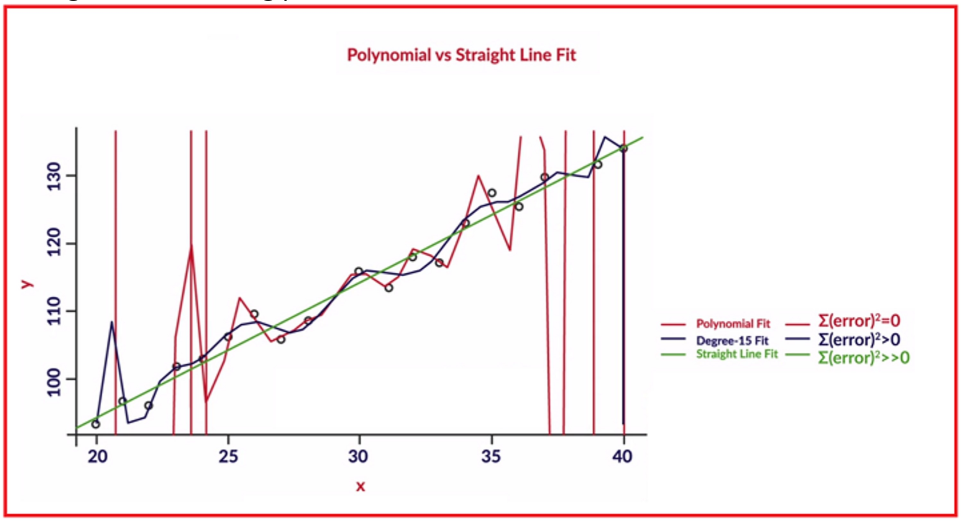
14.) Imagine a linear regression model, say car pricing dataset, where you consider transforming the features to log(x), sqrt(x) etc. Complexity here will be proportional to the number of coefficients, the magnitude of coefficients, degree of polynomial (x squared is a more complex feature than x) etc. In principle, you can take as many features as you like and transform them in complex ways to 'make a model which fits perfectly'. In short, you can increase complexity to reduce the error to zero. In the figure, however, the total error goes up (after the minima) as model compelxity increases. Why is that so?
By increasing complexity, I can reduce the training error, whereas the figure plots the test error
Feedback :*It is crucial to remember that we always talk about the test error (i.e. the expected error on a similar, unseen dataset) while discussing bias-variance tradeoff. If training error were concerned, we could just increase complexity to make the model pass through all the points.*

15.) Bias and Variance
The mental model models of the first and the second person, respectively, have (mark all that

apply):
Student 1 = high variance, student 2 = high bias
Student 1 = low bias, student 2 = low variance

16.) Model Variance - Regression
Rohit builds two linear regression models to solve the car pricing problem - model (1) has 3 features and model (2) has 11 features. Which model is likely to undergo a larger change when a new training dataset is used?
Model 2
Feedback :*Model 2 can change its 11 coefficients (and the constant term) to fit the new training data.*

17.)
Comprehension - Bias Variance Tradeoff
An artificially generated dataset was used to generate data of the form (x, 2x + 55 + e), where e is a normally distributed noise with mean zero and variance 1. Three regression models have been created to fit the data - linear,a degree-15 polynomial and a higher degree polynomial which passes through all the training points.



Polynomial vs Straight Line Fit

1.) Bias and Variance
The correct order of bias in the three models is:
Straight line > Degree-15 > Polynomial
Feedback :*The bias is high when the model is too simple.*

2.) Variance
The variance in the higher degree polynomial is said to be higher than the other two models because:
The model will change drastically from its current state if the current training data is altered
Feedback :*Variance refers to changes in the model as a whole when trained on a different dataset. Since the polynomial is trying to overfit the data, it will change drastically with respect to it.*

18.)Regularization
Regularization is typically performed while:
The learning algorithm uses the training data to produce a model
Feedback :*The learning algorithm performs the process while it is learning from training data.*

19.) Cross Validation
It is not often possible to use the validation set approach because:
The data available for training and testing is limited

20.) Training
Why should we have disjoint training and test datasets? i.e. model should not be tested with data it is trained on.
Suggested Answer
Any model has to be tested on how well it would work in the proverbial ‘real’ world. Because once a model has seen the data, it can pretty much attempt to ‘memorise’ it and once that is done, testing it on the same dataset will not help in determining its performance on unseen data. In an ideal scenario, when we have plenty of data, we should divide the data into three sets. First would be the training data. On which we shall train the model. . Second would be the validation data. On which we shall test the model and tune the hyperparameters. Third would be the test data, which we shall use for assessing our model.

21.) Variance
Which of the following is more likely to have low variance?
Weak Learner
Feedback :
Weak learners create simpler models which have lower variance. They are not able to model complex relationships and hence create a more generic model.

22.) Bias-Variance
Given two linear regression models on the same dataset with 100 attributes, model A has ten attributes while the model B ninety attributes, what can you say about the two models?
Model B has tried to memorize the data and when the training data changes a little, the expected results will change.
Feedback :
Model B has too many attributes and probably is overfitted.
Correct
The model B will have a much higher variance compared to model A.
Feedback :
Model B is memorising the data or overfitting and hence would have a higher variance.
Correct
Model A will have a higher bias
Feedback :
Model A has made a lot of assumptions about the data to keep the model simple leading to high bias.

23.) Bias-Variance
Consider that data is generated via a polynomial equation of degree 4 (i.e. the said polynomial equation will perfectly fit the given data). Which of the following statements are correct in this case?
Linear regression will have high bias and low variance
Feedback :
Linear regression would create a degree 1 polynomial which would be less complex than the degree four polynomial and hence would have a higher bias. Since the model is less complex and won’t overfit, it would have a low variance.
Polynomial equation of degree 4 will have low bias and Low variance
Feedback :
Since the equation fits the data perfectly, bias and variance will both be low here.