SuperDataScience

10 SAMPLE EXAM QUESTIONS
+ COMPLETE ANSWER KEY

MACHINE AND DEEP LEARNING BASICS — PART #1

1. **You have a build an image classification model and on training the model you find that training accuracy is 97 % and the validation accuracy is 65%. Why is this difference in the values?**

   A It may be due to the fact the model is underfitting the training data and fails to generalize

   B It may be due to the fact the model is overfitting the training data and fails to generalize

   C It may be due to the fact the model is overfitting the training data, though it generalizes well

   D It may be due to the fact the model is underfitting the training data, though it generalizes well

2. **You have been given access to student marks data and you want to predict the top performing students using AWS Linear Learner. On training you find that the model is overfitting the data, which metric be used to tune the model so as to avoid overfitting?**

   A Validation metric

   B Training metric

   C Overfitting metric

   D Underfitting metric

3. **Jack is tasked to train a regression model and after training he finds that training accuracy is around 96% and the validation accuracy is very low. What can he do to improve the validation accuracy?**

   A Increase the number of epochs and increase the dimensionality

   B Encode the data using one-hot encoding

C Train the model for longer period of time

D Add early stop to prevent overfitting and also try to add more data for training process

4. **Residual is the actual value minus the predicted value. On plotting the residual in the histogram, we observe the distribution is moved to the right of zero. What does it mean?**

A This means the model is overestimating

B This means the model is stuck at local minima

C This means the model is underestimating

D This means the model is stuck at global minima

5. **You are working on machine learning model and you have split the available data into 80% training set and the rest as test set. After training, you observe that the training accuracy is 97% and test accuracy is 56%. What is the best possible reason for this and how can you correct it?**

A It is because of the data imbalance between test and train set. You can randomize the data and then split

B Model is overfitting and we have to tune the parameters

C Use can use different model to this particular problem

D Model is underfitting and we have to tune the parameters

6. **You are building a regression model and after training you find that the model's training loss has remained the same or almost around the same value after some epochs. What might be the cause and how can you improve it?**

A Model might have reached the global minima and you can lower the learning rate

B Model might have reached the local minima and you can increase the learning rate

C Model might have reached the global minima and you can increase the learning rate

D Model might have reached the local minima and you can lower the learning rate

7. You are working on a Machine Learning regression problem and you are using RMSE as a metric for evaluating the model. After initial training value of RMSE was 4 and after some tuning RMSE value changed to 1.9. What does this mean?

A This means recall is reduced

B This means accuracy is increased

C This means precision is increased

D This means accuracy is decreased

8. John is building a time-series model and after training and validation, he finds that his model is overfitting the data. What changes can you make to the model architecture to avoid overfitting?

A Increase the epochs

B Use different activation functions

C Use early stopping

D Use bias-variance trade

9. What is the difference between positive number residual and negative number residual in regression?

A Positive number means model is underestimating and negative number means model is overestimating

B Positive number means model is overestimating and negative number means model is underestimating

C Positive number means model accuracy is high and negative number means model accuracy is low

D Positive number means model accuracy is low and negative number means model accuracy is high

10. You are tasked to build a ML model and you have been given access to the data. You want to split the data into training and testing data. Which is the most appropriate value for train-test split?

A 50 – 50

B 95 – 5

C 70 – 30

D 55 – 45

ANSWER KEY:

1. B	2. A	3. D	4. C	5. A
6. D	7. B	8. C	9. A	10. C