Quiz Submissions - Q3-1

Parsa Yadollahi (username: parsa.yadollahi@mail.mcgill.ca)

Attempt 1

Written: Jan 27, 2022 12:55 PM - Jan 27, 2022 12:57 PM

Submission View

Released: Jan 27, 2022 4:00 PM

Question 1 **1 / 1 point**

The reason behind curse of dimensionality is:

- ✔ The distances between any two data points become similar as the dimensions grows.
- Computing power in recent years is still not capable enough to handle large-scale computation on high-dimensional data.
- The data points get exponentially closer to the center of the input space as the dimensions grow.
- Variations in data will increase dramatically as data dimension goes up, so that a machine learning model will have harder time training on the data.

View Feedback

Question 2 **1 / 1 point**

You are performing 5-fold cross-validation to select the best hyperparameter values. You have two hyperparameters and you decide to perform a search over 10 possible values of each of the two hyperparameters and try all possible combinations (a.k.a grid search). How many times should you train your model to get a good estimate of the best hyperparameter values?

(For example, when you run a 5 fold cross validation, you are training the model 5 times.)

- 10
- 5
- 50

✔ 500

View Feedback

Question 3 **1 / 1 point**

In a perfect machine learning algorithm, the AUC (area under the curve) of an ROC (Receiver operating characteristic) curve will be

0

0.5

✔ 1

View Feedback

Question 4 **1 / 1 point**

For a cancer screening model for predicting whether a patient needs to undergo further tests and be seen by a doctor, if you have to choose between a model with high precision and a model with high recall, which one would you pick?

✔ Model with high Recall

Model with high Precision

View Feedback

Question 5 **1 / 1 point**

What's the definition of precision? (T: True, F: False, P: Positive, N: Negative, e.g., TP means True Positive)

TP / (TP + FN)

None of above

TP / (TP + TN)

✔ TP / (TP + FP)

Question 6 **1 / 1 point**

You are given four models with their corresponding number of parameters, training and validation accuracies (with standard deviation, denoted as stdev). Which model would you use?

- Model: 800 params, train accuracy: 95% (stdev: 2%), validation accuracy: 73% (stdev: 5%)
- Model: 80 params, train accuracy: 72% (stdev: 4%), validation accuracy: 66% (stdev: 5%)
- Model: 500 params, train accuracy: 88% (stdev: 12%), validation accuracy: 83% (stdev: 16%)
- ✔ Model: 300 params, train accuracy: 84% (stdev: 6%), validation accuracy: 81% (stdev: 5%)

View Feedback

Question 7 **1 / 1 point**

Why is there a difference between KNN model performance on image data such as MNIST and on random high-dimensional data?

- Images in MNIST dataset are not large enough.
- KNN is designed specifically for image classification, so it has strong power on classifying hand-written digits.
- MNSIT consists of single-channel image data.
- ✔ The pattern in pairwise distances for pixels from a hand-written digit in the MNIST dataset does not follow that of randomly generated data.

View Feedback

Attempt Score: 7 / 7 - 100 %

Overall Grade (highest attempt)**:** 7 / 7 - 100 %

Done