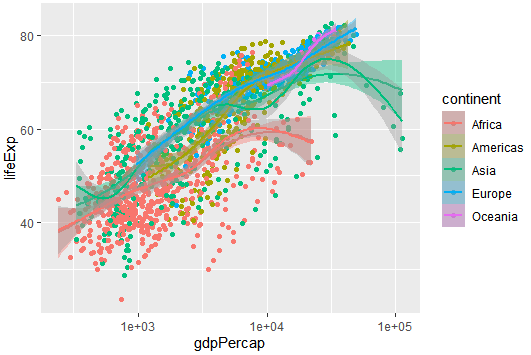
Lab 4

Group 8

**Exercise 1**

Below is the result of the code executed.



Here is the ranking of the models from most underfit to most overfit.

Europe

Americas

Oceania

Asia

Africa

1. Yes, the conclusion of the result changes. This is due to randomization of folds and variation in the number of seeds used.

Question 3

1 Africa

[ [118 45]

[47 15]]

America

[ [121 43]

[50 17]]

Europe

[ [125 49]

[32 12]]

Asia

[ [100 30]

[40 24]]

2

Africa is the best model.

Model Justification

a) overall accuracy

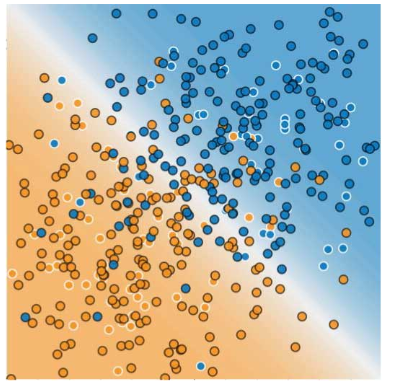
The variance of the prediction error is a good tool that helps us to determine and conclude that this model is our best model. It is noticed that the model omits no variables leading to formation of the overfit model. Due to this, the model is termed to be accurate and the best model to be used compared to other models.

b) Errors

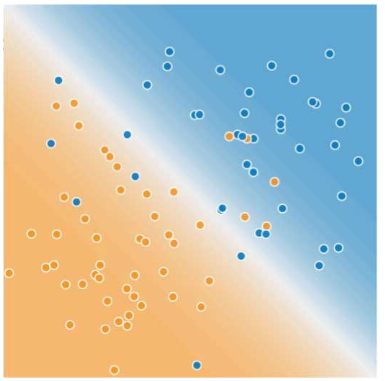
Basing on the errors of models, it is noticed that most of the errors of this Africa model are better than the worst ones. The specificity and sensitivity of Africa model is high compared to other models hence ranking this model as the best model of all the present model.

**Exercise 2**

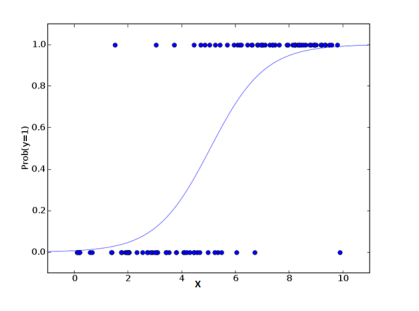
Training data



Test data



2



3.

Probability has always had a positive value since it never goes below zero value as shown in the graph above. Due to this the probability of having a y for yes value is very high.

Confusion matrix for the above model

[ [40 20]

[2 10]]