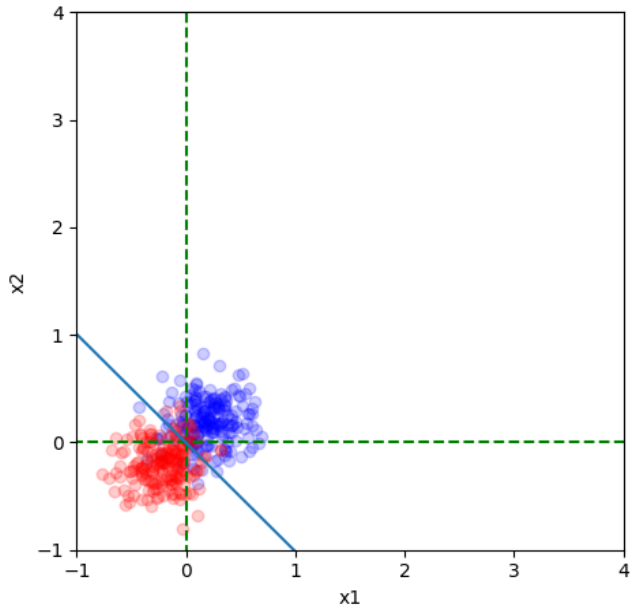
[CMPUT 466/566] Machine learning

Coding Assignment 2

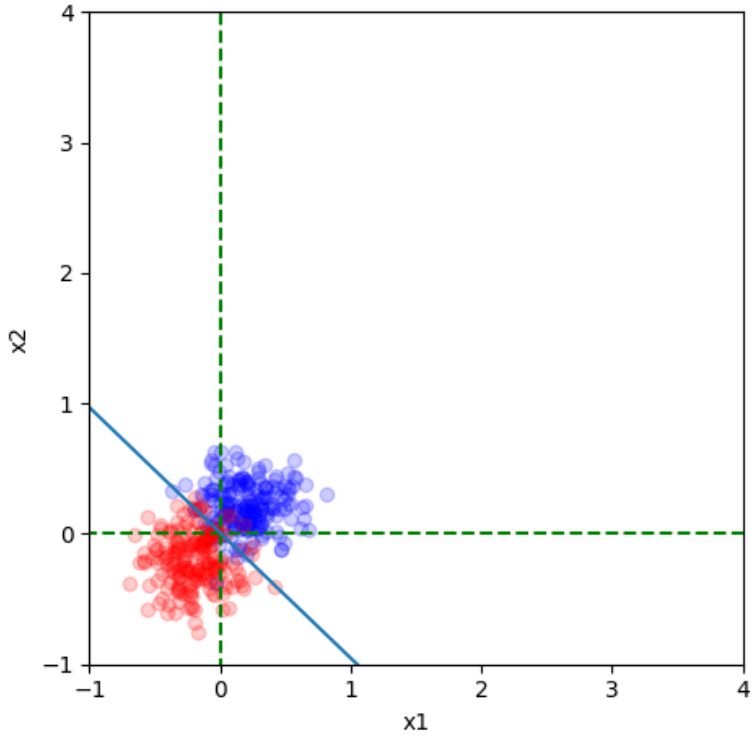
Question1:

Accuracy of linear regression on dataset A: **0.91**
Accuracy of logistic regression on dataset A: **0.925**
Accuracy of linear regression on dataset B: **0.75**
Accuracy of logistic regression on dataset B: **0.9425**

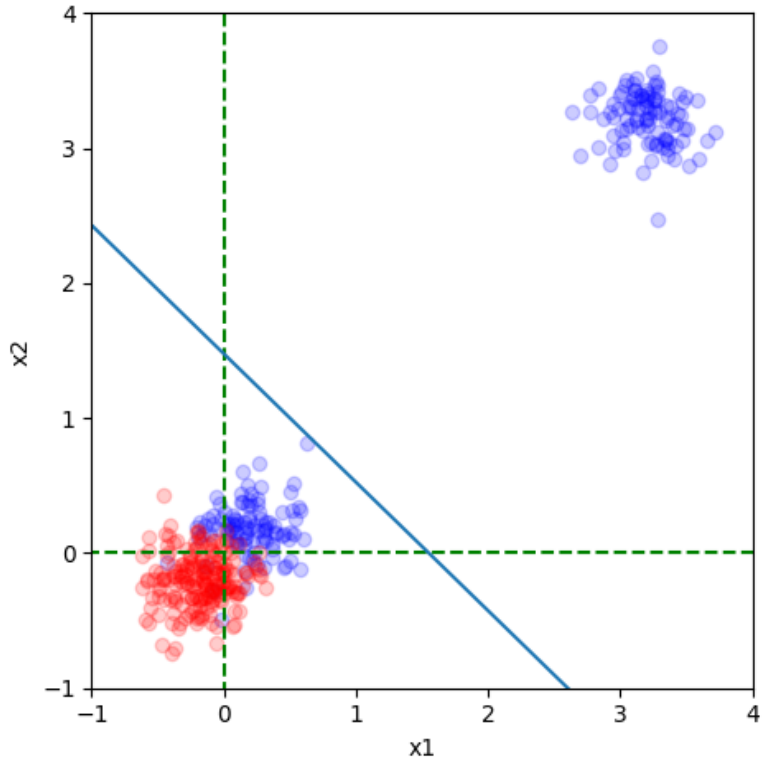
Data_A_Linear_Regression:



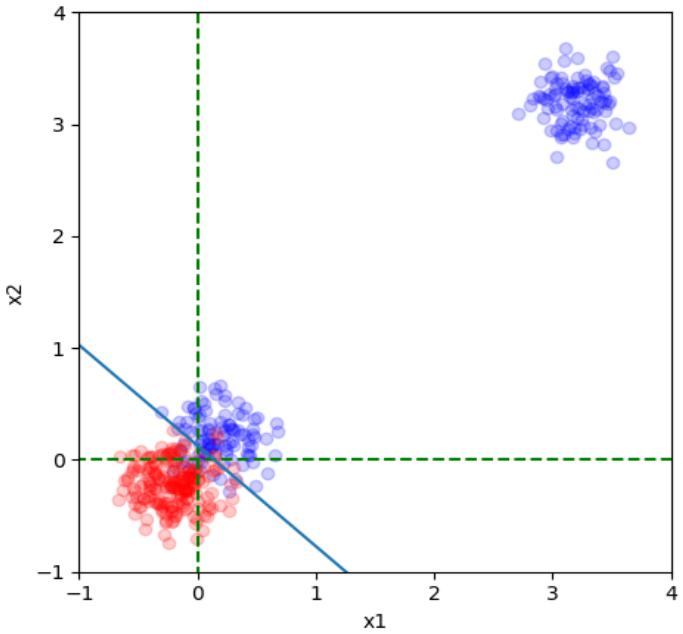
Data_A_Logestic_Regression:



Data_B_Linear_Regression:

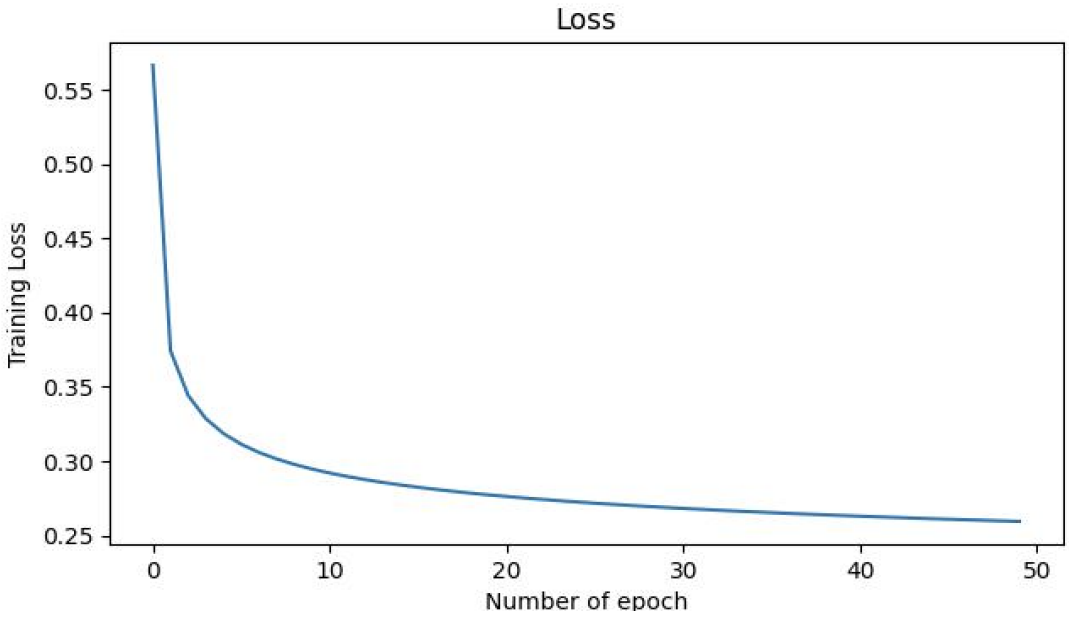


Data_B_Logestic_Regression:

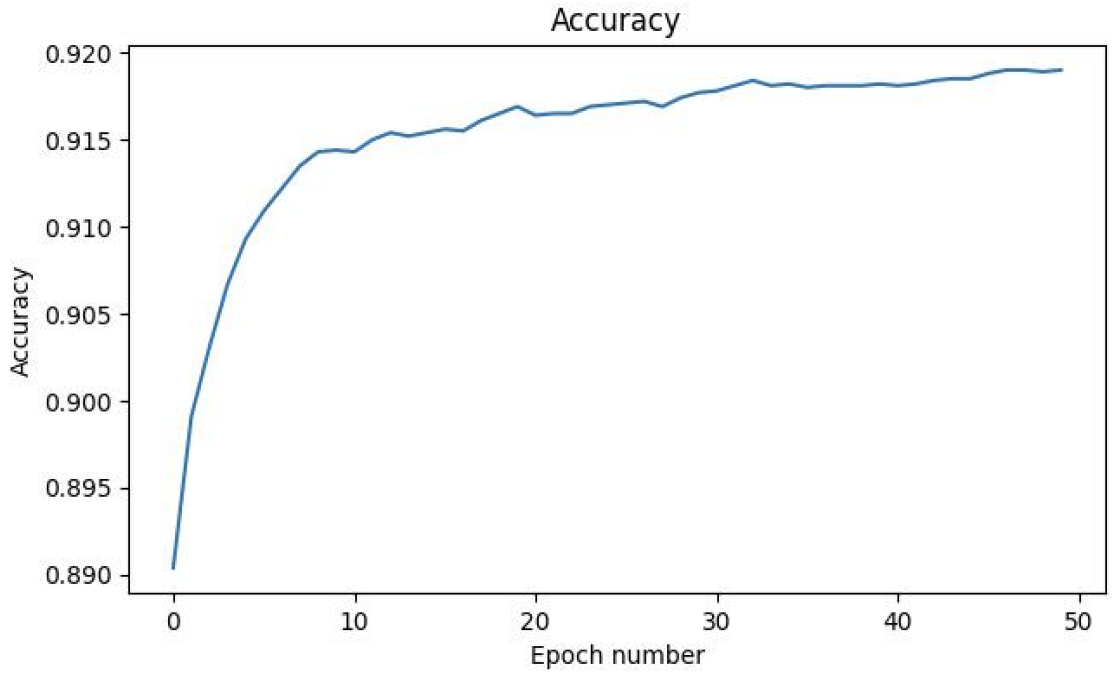


Question2:

1. The number of epoch that yields the best validation performance is **49**
2. The **validation performanc**e (accuracy) in that epoch is **0.919**
3. The **test performance** (accuracy) in that epoch is **0.9233**
4. The learning curve of the training cross-entropy loss,



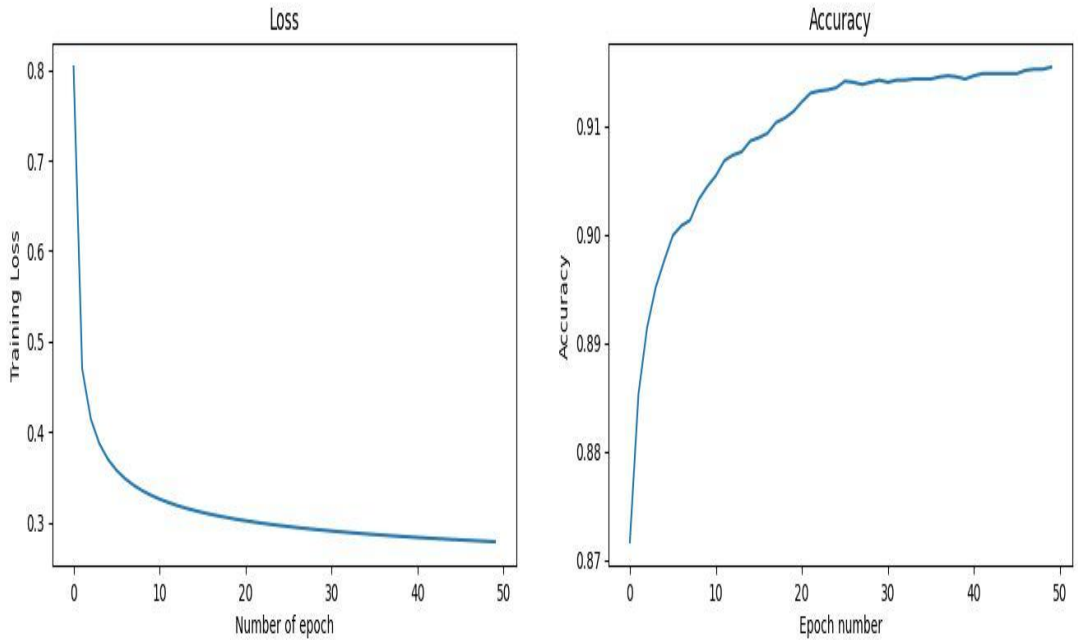
The learning curve of the validation accuracy:



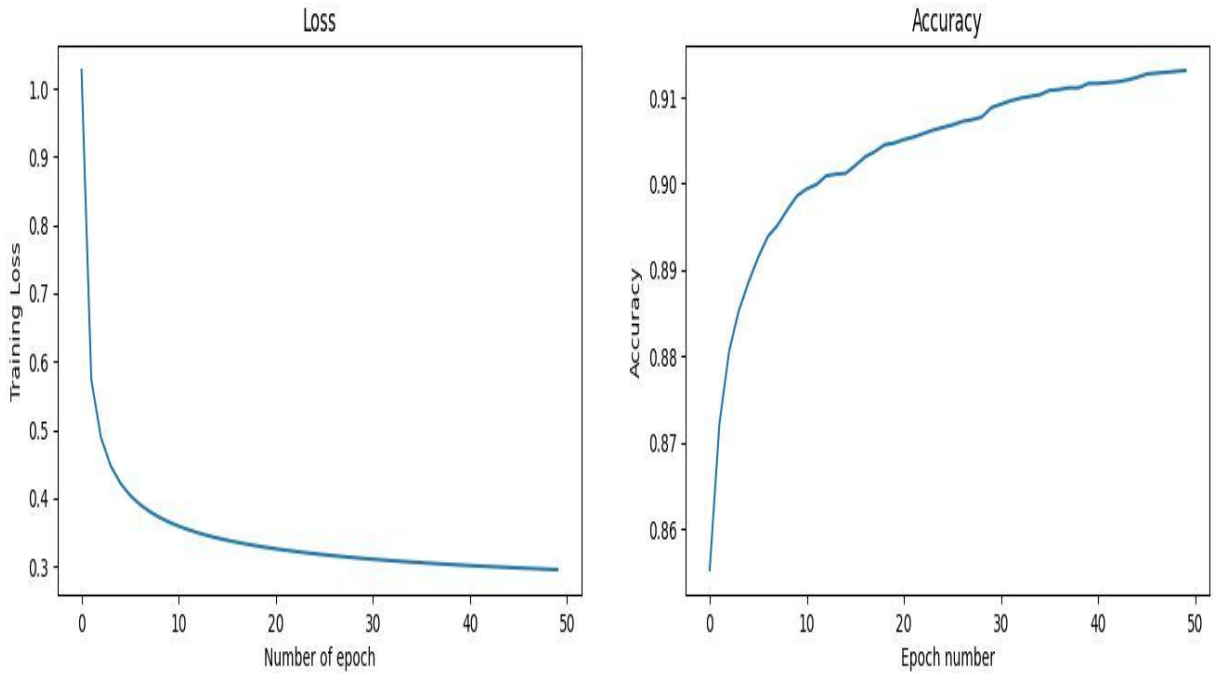
Scientific Question:

How can batch size influence the overall performance of the model?Will a model with larger batch size lead to a better performance?

Without changing other hyperparameters such as alpha and max epoch, increase the batch size from to 100 to 300 yield the following results:
Test performance: 0.9209
Validation performance: 0.9154



When change the batch size from 300 to 600:
Best epoch 49
Test performance: 0.9193
Validation performance: 0.9131



Conclusion: With a larger batch size, the number of test accuracy increases slightly larger whereas the number of validation accuracy drops. Therefore, increasing the batch size for the model will not enhance the performance.