UNIVERSITY 4 INDUSTRY

Example Solution
The Taste of Machine Learning

You can compare your answers to the example solution below.

Question 1: Which two qualities of wine are most frequent in the data set?

Answer 1:
- Wines of quality 5 and 6

Question 2: What are good reasons to replace the variable 'quality' by the variable 'review'? Which impact will this have on our machine learning model?

Answer 2:
- It is more difficult to predict 10 different classes than 3.
- We do not have so much data (only 160 examples per class in average).
- This might sound like a good idea to make the accuracy of the model better, but in the end, it will not help us to create a trustworthy model.

Question 3: The wine of which review occurs most frequently in the data set?

Answer 3:
- Review 2

Question 4: Which accuracy did your model achieve?

Answer 4:
- ~98/99%

Question 5: What does the accuracy score tell you about the quality of your model?
Hint: Please describe the drawbacks of using accuracy as a measure to judge the quality of a model.

Answer 5:
- The accuracy is high, but this does not automatically mean that the model is good. The accuracy just tells how many wine qualities were predicted correctly
- It does not take into account how the quality of wines is distributed. The accuracy does not directly show whether outliers are predicted correctly

Question 6: Please interpret the confusion matrix of you model. What does the outcome tell you about the accuracy score of your model and the accuracy score in general?

Answer 6:
- The confusion matrix tells us that all wines have simply been predicted to have review 2.
- Since 1571 out of 1599 wines got "review two", the model achieves a very high accuracy of over 98%.
- Of course, this means that the model is completely useless. Only if our model achieves an accuracy score of above 98,25%, it means it does a better job than putting everything into review category 2.

Question 7: Share some ideas with us. How could a better model be created? How would you proceed with the wine quality prediction task?
Please name at least two ideas.

Answer 7:

- Find more training examples
- Change the categorization to include more examples in category 1 and 3
- Make the quality increments course again
- Try a different machine learning model
- Fine-tune the machine learning model