

MLCourse-LU

s2538334

Assignment 1

1. One way to improve the experiment is to change the postive class from "virginica" to "setosa".
 2. Another way to improve the experiment is to use a linear kernal instead of "rbf".
 3. A third way to improve the experiment is to increase the amount of test samples; increasing the fraction to 0.6 yields slightly better results.
-
4. The SVM, as that one seems to have a much higher true positive rate as opposed to the false positive rate.
 5. This is because the data regenerates upon each re-run, and therefore, as the model is dependent on the data, the model also changes. A way to prevent that is to settle upon one seed with the data regeneration, but this is not a good idea since then fluctuations in the model cannot be seen, and potential outliers can be accepted as a regular result.
 6. The false positive rate increases more than the true positive rate, so your model is unreliable.
-
7. For both SVM and RandomForest this seems to be class 8, as the AUC value is the lowest in both cases. The curve for d=8 seems to cover the least area, and as AUC is a measure for 'how correct' a model is, the higher this value, the easier a model is to learn. For d=8 this value is the lowest, and therefore this is the hardest class to learn for both models.
 8. It could be because the 8 as a number looks a lot like other digits (such as 0, 4, 6, 9). This leads to worse distinguishment by the model relative to other digits and therefore also a worse model performance.
-
9. 15, because the true positive rate is higher than at point 14, while the false positive rate is the same, yielding a better model at point 15.
 10. 15, because the true positive rate is the same at both points, while the false positive rate is higher at 16.
 11. Illness screening, as the increase in false positive rate does not matter as much as the increase in false negative rate, for example.
 12. Fraud screening, as you need the false positive rate to stay as low as possible.
 13. It means that this point is somehow worse than assigning a value to a random class, i.e. it is more likely to make a false prediction than a correct one.