

Assignment 1 Elin Kile

GitHub link: https://github.com/KileElin/FYS2021_assignments

1a)

232725 rows = 232725 songs or samples.

18 columns = 18 features

1b)

Label 1 = Pop, 9386 samples

Label 0 = Classical, 9256 samples

1c) Done. See in the script.

1d)

I would say it depends. There are most overlap on the “liveness”, and only some overlap on the “loudness”. I believe just from looking at the plot that it will be possible to achieve over 80% accuracy.

2a)

For all values tested the cost decreased to a certain point and maintained that level of cost no matter how many more epochs you run. The higher the learning rate, the sooner the curve levels off. Above a certain learning rate of approximately 23, the loss curve started oscillating around a higher mean, indicating that the steps in the gradient descent were too big. If you go towards the other end, with smaller learning rates, the number of epochs needed to reach the levelling off of the curve increased.

2b)

When the learning rate was 0.1 and the model was trained with 1000 epochs, 10 times, the average accuracy of the test set was 92.88% (± 0.03), using t-statistics it was determined to be a significant difference between this and the accuracy of the training set which had an average accuracy of 92.56% (± 0.01), under the same conditions. While the difference is significant, it is not large. In this case I got a higher accuracy for the test set, which is surprising, since it should perform better on known data, than on new data? I will admit I don't fully understand why this is.

2c) Bonus. Was unable to do this.

3a)

Confusion matrix

TP = 1638	FP = 213
FN = 82	TN = 1795

3b)

The confusion matrix tells us more about the specificity and sensibility of the test, since we can see if the reduced accuracy (compared to 100%) is more due to false negatives or false positives. In this particular case it seems that false positives are more likely than false negatives, which means that more classical songs are classified as pop, than the other way around. The accuracy only tells us how accurate the predictions are over all.

3c)

Liveness is not a good measure, but loudness seems to be a better way to classify the different songs. There are however a lot of overlap, so classical songs with high loudness will likely be wrongly classified as pop.

As for suggesting a classical song that a pop fan would like. Looking at the plot of the pop and classical music songs, it looks like the loudness could be a good way to find a classical song that a pop fan would like. Sorting for loudness in the classical songs gave an interesting result, as the top five songs were more in the rock/metal genre than anything else. So, I checked the csv file and found that a lot of the classical songs were actually rock/metal. But as I assume this is not an oversight but rather a way for us students to prove that we actually did the work. Therefore, the song I will suggest is "Go faster" by Richie Kotzen.