

Study doc: regularisation experiment

how does the regularisation factors effect the embedding space

Fabian Gröger
fabian.groeger@stud.hslu.ch

Wednesday 13th May, 2020

Abstract

The experiment aims to show how the embedding space will change for different regularisation factors.

1 Introduction

The purpose of regularisation is that it prevents models from overfitting, by penalising high weights. The regularisation is calculated for all layers weights separately and is then added to the standard loss function, in this case, the triplet loss. The combined loss value is then used to calculate the gradients and update the weights of the model. There are a lot of different regularisation techniques, which successfully reduce the chance of overfitting. In this experiment, the **L2-Regularisation** is used and evaluated with different values. The parameter which will control the value of the **L2-Regularisation** is denoted as λ .

2 Hyperparameters

The hyperparameters used for this experiment are shown in table 1. The experiment will be conducted using a state of the art ResNet18 architecture on the DCASE dataset. The hyperparameters in section *Feature representation* as well as the sample rate are the default ones proposed by the organisers of the DCASE challenge within the baseline project. The regularisation factor λ will be evaluated for three different values [1.0, 0.1, 0.01].

Table 1: Hyperparameters used for the experiment

Hyperparameter	value
Dataset	DCASE
Model	ResNet18
Epochs	30-50
Batch size	128
Optimizer	Adam
Learning rate	1e-5
Margin	1.0
L2 regularisation factor	[1.0, 0.1, 0.01]
Embedding dimension	64
Prefetch batches	Autotune (-1)
Random selection buffer	64
Shuffle dataset	True
Random seed	1234
<i>Multi threading</i>	
Number of generators	16
Number of parallel calls	16
<i>Audio sample</i>	
Sample rate	16000
Sample size	10
Sample tile size	5
Sample tile range	5
Convert to mono	True
<i>Feature representation</i>	
Feature extractor	LogMelExtractor
Frame length	480
Frame step	160
FFT size	1024
Number of Mel bins	128
Number of MFCC bins	13

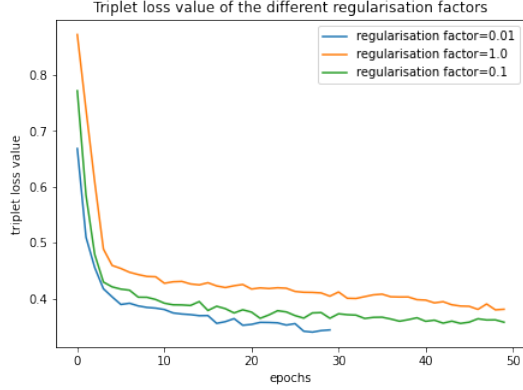


Figure 1: Plot of the triplet loss values of the different regularisation factors λ

3 Results

Comparing the effect of different regularisation factors on the embedding space is pretty hard, since comparing embedding spaces is not very straight forward and is a rather tricky task. This is mainly because of the fact, that to visualise the high dimensional embedding space in a way humans can perceive it, it has to be reduced to two or three dimensions. Therefore the original space can not be examined, and the visual representation is always an approximation of the space in a lower dimension. To still compare the embedding spaces and the effect of λ , a simple logistic classifier is trained on top of the resulting embedding spaces, which aims to show how well a simple classifier works with the embedding space. This will show how good the resulting embedding space is. The classifier is trained for 20 epochs with the same parameters as the embedding model (table 1).

Figure 1 shows the difference between the triplet loss values of the embedding model when changing the regularisation factor. It shows that the model with $\lambda = 1.0$ has the highest triplet loss value, which is obvious, since the bigger the regularisation factor is, the longer the model optimises the weights to satisfy the constraint of having very small weights. The lower λ , the faster the model optimises the triplet loss value. $\lambda = 0.01$ has the lowest triplet loss value.

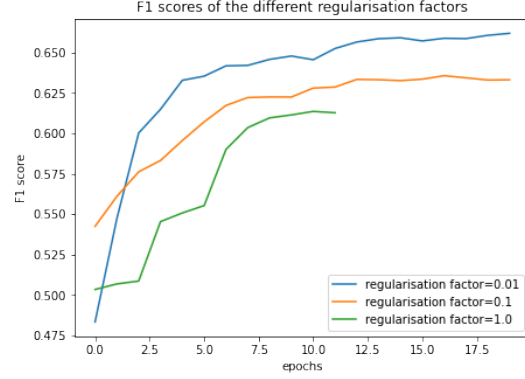


Figure 2: Line plot of the F1 score from the different classifiers trained on top of the resulting embedding spaces

Figure 2 shows the different F1 scores of the logistic classifier, which is trained on top of the embedding architecture. This provides an idea of how well the embedding space separates the classes and therefore gives a performance gain. The figure 2 shows, that the regularisation factor $\lambda = 0.01$ has the highest F1 score. This result is significant since this particular embedding space was only trained for 30 epochs, unlike the other models, who are trained for 50.

The figure 3 shows the corresponding categorical cross-entropy loss values of the trained classifiers. It shows that the loss of $\lambda = 1.0$ converges before $\lambda = 0.1$ or $\lambda = 0.01$. The figure further shows, that the loss values of $\lambda = 0.1$ and $\lambda = 0.01$ are not converged and would benefit from a longer training.

4 Conclusion

Figure 2 shows that the highest regularisation factor $\lambda = 1.0$ can be omitted since it did not accomplish an equally good F1 score. This result is expectable since a high regularisation factor forces the embedding model to work with very small weights and can, therefore, lead to performance loss.

The regularisation factor $\lambda = 0.1$ and $\lambda = 0.01$ only show very little difference in F1 performance as well as in triplet loss value. Never

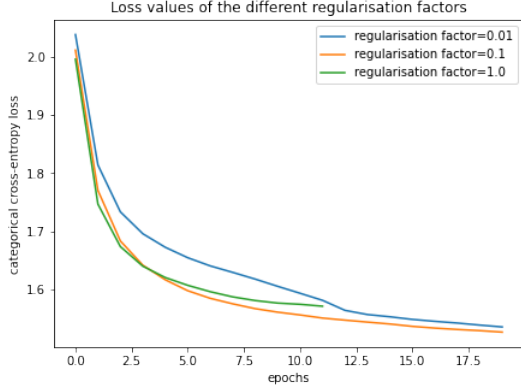


Figure 3: Line plot of the sparse categorical cross-entropy loss of the embedding size 16 and 64

the less, the factor $\lambda = 0.01$ provides slightly better results than $\lambda = 0.1$.

The triplet loss value and the loss value of the classifier of the regularisation factor $\lambda = 0.01$ is still decreasing at the end of the training, which indicates that the model can benefit from further training. Further training will increase the resulting embedding space and the F1 score of the classifier until it converges.

The figure 1 and 2 further indicate one very important conclusion, that the value of the triplet loss is directly related to how well the classifier performs, which is further an indication how well the created embedding space works. Therefore the assumption is, that the lower the triplet loss value, the better the classifier performs and the better the embedding space is.

The experiment shows that the regularisation factor $\lambda = 0.01$ gives the best results out of the three since it has an optimal triplet loss value, a high enough F1 score and the loss value still decreases after 20 epochs of training the classifier.

5 Next steps

The experiment showed that the regularisation factor λ has an essential impact on the embedding space and should therefore not be chosen to be too big nor too small since this either leads to over- or underfitting. The optimal

regularisation parameter λ should be further examined in later and more extensive experiments with a lot more values for λ .

For the next experiments, the regularisation factor $\lambda = 0.01$ is chosen.