Homework 2

Foundations of Computer Vision

Question 1:

Confusion matrix is used for evaluation of machine learning models allowing us to visualize the performance of the model.

		Actual class	
		P	**N**
Predicted class	**P**	**TP**	FP
	N	FN	**TN**

True positives (TP): Predicted as part of the class, that are actually part of the class.

True negatives (TN): Predicted as not a part of the class, that are not part of the class.

False positives (FP): Predicted as part of the class, that are actually not part of the class.

False negatives (FN): Predicted as part of the class, that are actually part of the class.

Confusion matrices for each label:

```
[[221   4]
 [ 22   3]] 01beach
[[208  17]
 [ 13  12]] 02forest
[[205  20]
 [ 18   7]] 03mountain
[[210  15]
 [ 19   6]] 04city
[[172  53]
 [ 13  12]] 05suburb
[[202  23]
 [ 12  13]] 06street
[[214  11]
 [ 19   6]] 07bedroom
[[217   8]
 [ 24   1]] 08kitchen
[[217   8]
 [ 23   2]] 09livingroom
[[204  21]
 [ 17   8]] 10store
```

Accuracy tells you how "right" your predictions are. It is the sum of true positives and negatives divided by the total data in datasets. One issue with accuracy is that it assumes positive and negative errors are equal. In addition, depending on the problem you're trying to solve you may be concerned about minimizing the prevalence of false positives over false negatives or vice versa. The next measures help you deal with this issue.

$$Accuracy = \frac{TruePositives + TrueNegatives}{TotalData}$$

Recall tells you how often the model chooses the positive class when the observation is in fact in the positive class. It is calculated by dividing the number of true positives in the matrix by the total number of real positives in the data.

$$Recall = \frac{TruePositives}{TruePostives + FalseNegatives}$$

Precision measures how often a model is correct when it predicts the positive class. It is calculated by dividing the number of true positives in the matrix by the total number of predicted positives.

$$Precision = \frac{TruePositives}{TruePostives + FalsePositives}$$

F_1 score is the harmonic mean of the precision and recall. For evaluating the F-score for multiclass problems it is done by micro-averaging (biased by class frequency) or macro-averaging (taking all classes as equally important). In our report we have used macro-averaging, different formulas can be used: the F-score of (arithmetic) class-wise precision and recall means or the arithmetic mean of class-wise F-scores.

$$Recall = \frac{Recall * Precision}{Recall + Precision}$$

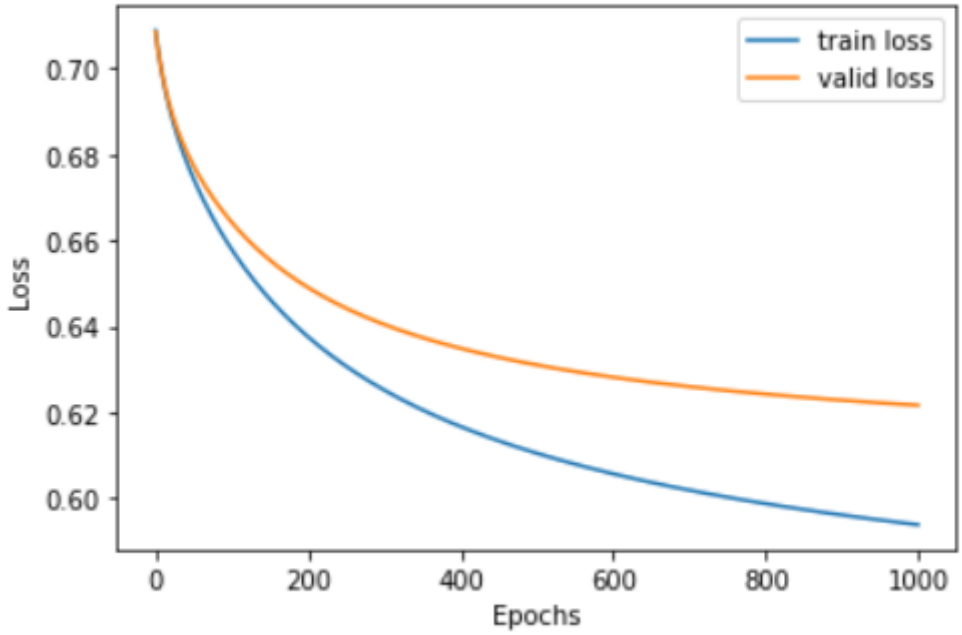
Performance of the model on each class:

	precision	recall	f1-score
01beach	0.43	0.12	0.19
02forest	0.41	0.48	0.44
03mountain	0.26	0.28	0.27
04city	0.29	0.24	0.26
05suburb	0.18	0.48	0.27
06street	0.36	0.52	0.43
07bedroom	0.35	0.24	0.29
08kitchen	0.11	0.04	0.06
09livingroom	0.20	0.08	0.11
10store	0.28	0.32	0.30
accuracy			0.28
macro avg	0.29	0.28	0.26
weighted avg	0.29	0.28	0.26

Question 2: Tested for 1000 epochs

The following image shows the results of this question before adding a regularization factor.

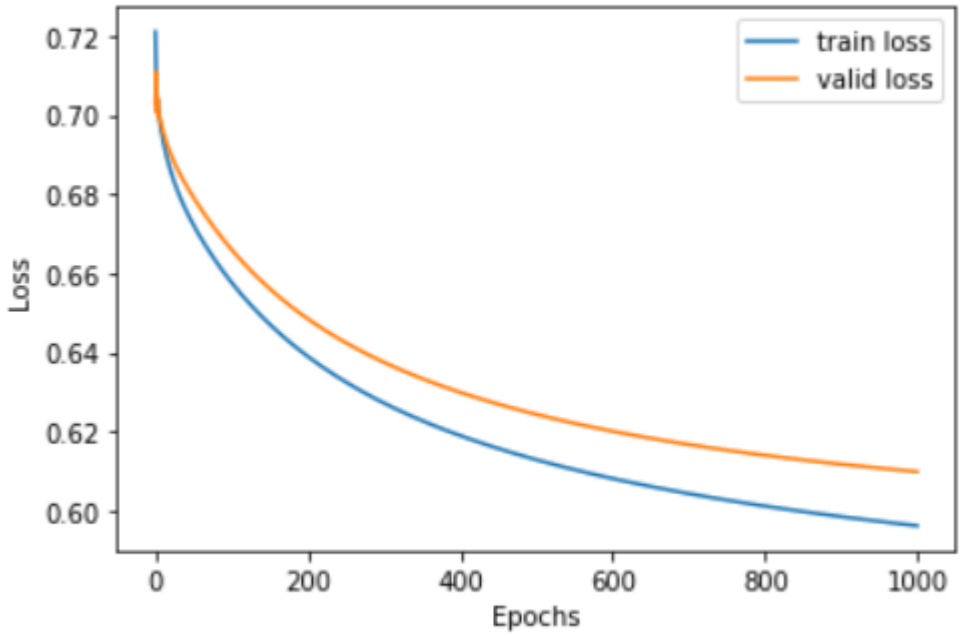
```
best_validation_error: 0.6216512117752823
```




```
Accuracy of test set is:  0.709
```

The following image shows the results of this question after adding a regularization factor.

```
best_validation_error: 0.6099523058101498
```




```
Accuracy of test set is:  0.7076666666666667
```