Assignement 1: Report bonus

a) Improvements

	lambda	Nr epochs	nr batch	eta	acc_train	acc_valid	acc_test
Basic	0	40	100	0.001	0.4548	0.3864	0.3895
Basic	0.1	40	100	0.001	0.4478	0.3895	0.3961
1) Shuffling	0	40	100	0.001	0.4531	0.3828	0.3858
1) Shuffling	0.1	40	100	0.001	0.4469	0.3867	0.3917
2) More data	0.1	40	100	0.001	0.42247	0.409	0.4122
3) Grid Search	0.1	40	80	0.002	0.4228	0.409	0.4081
4) Ensamble	0.1	40	100	0.001	-	-	0.4111
5) Decaying learning rate	0.1	40	100	0.001	0.42433	0.406	0.4104

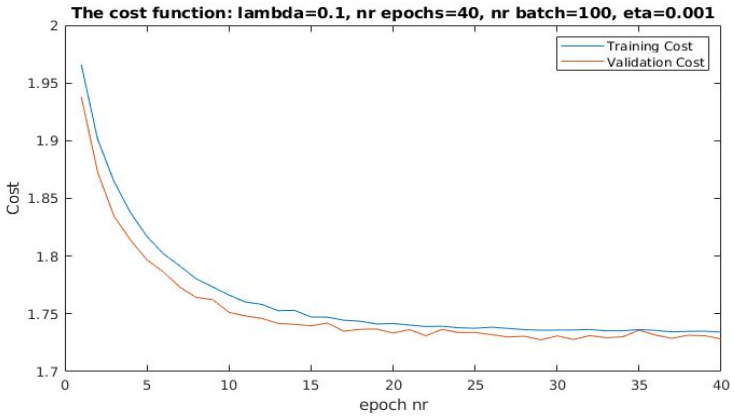

Fig1. Cost function using all available training data

From the 5 improvements presented in the table above the ones that brought the largest gains are:

- Using more data
- Ensamble
- Decaying learning rate

b) SVM(hinge) multi-class loss

	lambda	Nr epochs	nr batch	eta	acc_train	acc_valid	acc_test
1) Cross Entr	0	40	100	0.1	0.3819	0.2536	0.2496
1) SVM	0	40	100	0.1	0.2808"	0.1721	0.1713
2) Cross Entr	0	40	100	0.001	0.4576	0.3802	0.3903
2) SVM	0	40	100	0.001	0.4259	0.3456	0.3435
3) Cross Entr	0.1	40	100	0.001	0.4459	0.3893	0.3962
3) SVM	0.1	40	80	0.002	0.4333	0.3568	0.3522
4) Cross Entr	0.1	40	80	0.002	0.3962	0.3668	0.3738
4) SVM	0.1	40	100	0.001	0.4011	0.367	0.362
5) Cross Entr (Full data)	0.1	40	80	0.002	0.423	0.402	0.4114
5) SVM (Full data)	0.1	40	100	0.001	0.33257	0.326	0.3099
6) Cross Entr (decay)	0.1	40	100	0.001	0.42433	0.4104	0.406
6) SVM (decay)	0.1	40	100	0.001	0.41441	0.399	0.3897

The numbers of the figures below correspond to the numbers on the first column of the table above.

-

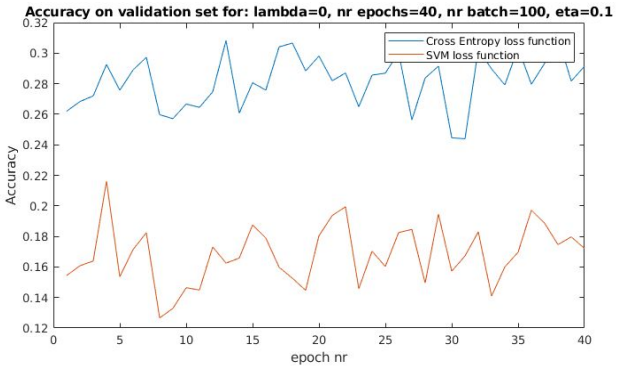

Fig. 1

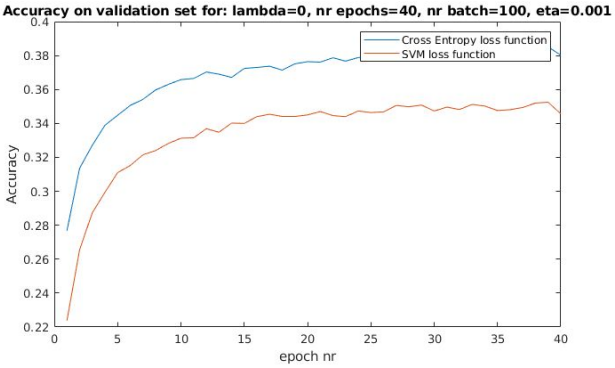

Fig. 2

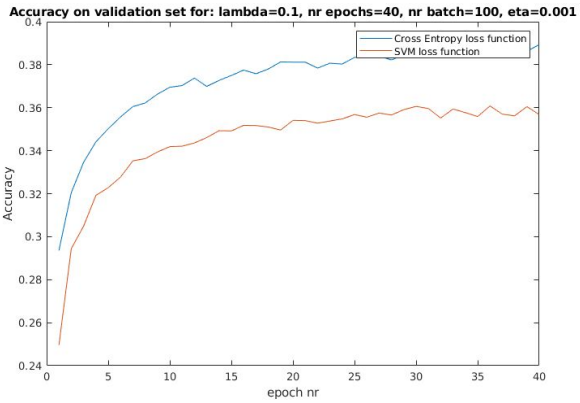

Fig. 3

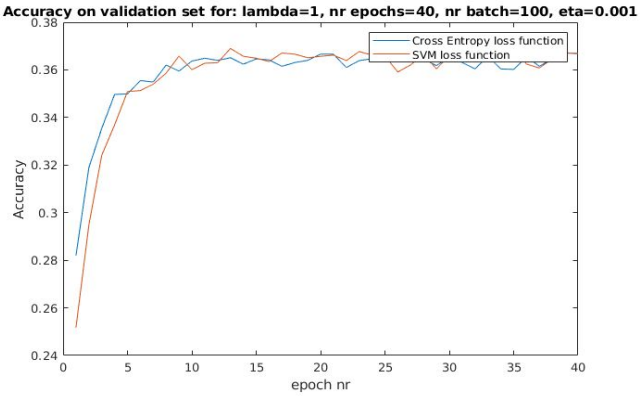

Fig. 4

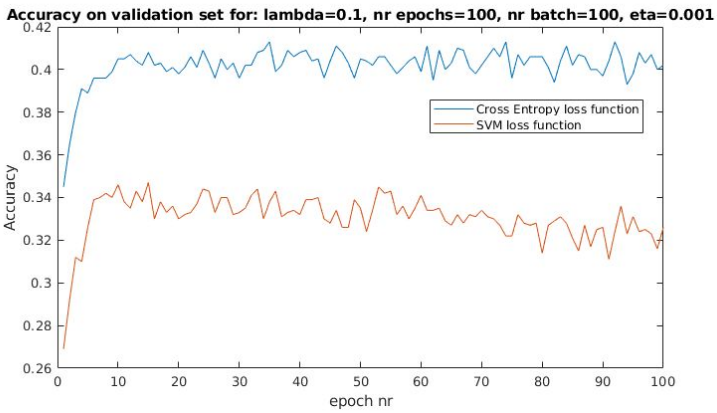

Fig. 5 Accuracy for complete training set

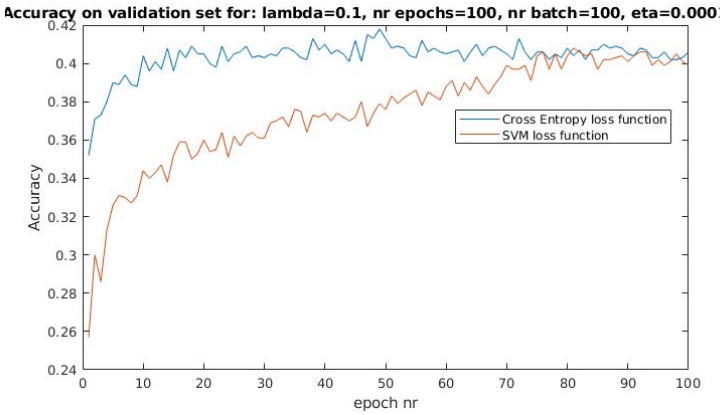

***Fig. 6* Accuracy for decaying learning rate**

Conclusion: In general, using Cross-Entropy as a loss function tends to deliver better accuracy than using the SVM(hinge) loss function. But we notice that once we introduce decaying learning rate both methods perform the same, with the SVM loss being a little slower in the first epochs.