Teachers: J. Hennebert, M. Melchior
Assistants: Y.-I. Beffa, S. Hackstein, O. Zayene

Practical Work 04 – 17/03/2022
Model Selection

Objectives

The main objectives of this Practical Work for Week 4 are the following :

a) Play through an example of overfitting and determine the optimal model complexity.

b) Perform a hyper-parameter tuning of an MLP with dropout regularisation for the Fashion-MNIST dataset.

c) Implement 5-fold cross validation to train/validate an MLP for Fashion MNIST and provide confidence limits for the validation accuracy per epoch.

Submission

— **Deadline** : Wednesday 30 March, noon

— **Format** : Jupyter notebooks including comments and results. For the solution of Exercises 2-4 you can prepare one single notebook. Clearly indicate which section relates to what exercise.

— Please **only one submission per group** - with a clear indication of the group id.

Exercise 1 Optional : Bias-Variance Tradeoff

The objective of this exercise is to build a classification system to predict whether a student gets admitted into a University or not based on their results on two exams [1].

You have historical data from previous applicants that you can use as a training set. For each training example i, you have the applicant's scores on two exams $(x_1^{(i)}, x_2^{(i)})$ and the admissions decision $y^{(i)}$. Your task is to build a classification model that estimates an applicant's probability of admission based on the scores from those two exams.

In the notebook `overfitting_polynomial2d_stud.ipynb`, you'll find the code to load the data and further instructions.

a) Construct different (polynomial) models of different complexities (different degree related to the number of parameters). Train these models with the training set and use the trained parameters for doing predictions. Measure the error rate on the training and the test set.

b) Determine the model best suited for the problem at hand and justify why it is the best model. Prepare a suitable summary plot with the train and test error vs polynomial degree.

Use the data `scores_train_1.csv` and `scores_test_1.csv` for training and testing, respectively.

Exercise 2 MLP Parameter Tuning with Regularisation

Carry through a model selection procedure for a classifier (based on MLPs) for the Fashion-MNIST dataset.

As a baseline use one hidden layer with 200 units, batch size 64, no regularisation. For this baseline, identify a reasonable learning rate and a reasonable number of epochs.

Explore along the following dimensions (by always starting from the same baseline) :

— model complexity : hidden units in the first layer - 5 different number of units

— regularisation : dropout rate for hidden layers and input layer - 5 different settings.

For each setting above, properly tune the learning rates and number of epochs. Use accuracy as evaluation metric.

For a given dropout rate, do you need to increase the model complexity ? If yes - how much ?

In the file `FashionMNIST_mlp_stud.ipynb`, you find the functionality to load the data and the skeleton for how to proceed.

What are your favourite model settings ?

1. Data source : Andrew Ng - Machine Learning class Stanford

Exercise 3 5-Fold Cross Validation

For the your favourite model in Exercise 2 above, carry through 5-fold cross validation.

Remember the per epoch validation accuracy for each fold and prepare a plot with the confidence limits (as shown in the lecture).

Given the information about these confidence limits does this has an impact on your judgement about your favourite model identified in the previous exercise ?

Exercise 4 Confusion Matrix

For the your favourite model in Exercise 2, compute the confusion matrix.

Compute the confusion matrix at the end of each epoch and analyse how it changes over the course of the training.

Identify which clothes are more difficulties for the model to be classified, i.e. which are confused.

Finally, also determine

a) overall and class accuracy

b) overall and class recall

c) overall and class precision

What are the worst and best classes in terms of precision and sensitivity (recall) ?

If you prefer you can alternatively use MNIST digits.

Exercise 5 Optional : Review Questions

a) Explain the terms bias and variance. What are the factors that make the bias larger or how can it be made smaller ? What factors lead to a large variance or how can it be reduced ?

b) Why is the training error an increasing function of the split ratio (fraction of samples used for training) ? Why is the validation error a decreasing function of the split ratio ?

c) Describe in words how you construct a confusion matrix. How can you compute the accuracy from it ? How does this relate to the error rate ?

d) Describe the terms class accuracy, recall and precision. Describe typical situations where you would like to obtain a high recall or a high precision, respectively.