Exercise Sheet 7

Regularization

Deadline: 12.01.2021, 23:59

Exercises

Exercise 7.1 - Norm penalty   (0.5 + 0.5 + 0.5 = 1.5 points)

Read Chapter 7 on Regularization from Deep Learning book.

a) Why do we penalize only weights, but not biases? Explain in your own words.

b) Let's consider a case of fitting linear regression to our Titanic dataset from previous exercise: we want to predict the *price of the ticket*. As the input we use *passenger class* and *age*. As we found out, *age* does not provide much information for our model. If we apply L2 norm penalty, which weight do we expect to be penalized more? Why? Think about the curvature of the loss function.

c) Explain, why L1 norm is used as a feature extraction mechanism. Support your answer with an illustration of the effect of L1 (similar to Figure 7.1). What effect will L1 norm regularization have on the model defined above?

Exercise 7.2 - Dataset Augmentation   (0.5 + 1 + 0.5 + 0.5 = 2.5 points)

Read this article by Shahul Es and about data augmentation in natural language processing. For a fuller picture, you can also have a look at this article by Amit Chaudhary. Answer the following questions:

a) In computer vision, data augmentation happens on the go, whereas in NLP data is augmented before training. Why is it so? Explain in 3-4 sentences.

b) Data augmentation in NLP is very task-specific and should be applied carefully. Provide an example of a augmentation method & task pair where data augmentation might harm the model. Justify your answer (2-4 sentences). A list of NLP tasks you can find here (you can also give an example with a task not from this list).

c) If we perform K-fold cross-validation, do we augment data before or after splitting the data? Why? Give at least two reasons.

d) Does the proportion of augmented data depend on the size of the training dataset? Explain why. Hint: think about the relation between training dataset size and overfitting.

Exercise 7.3 - Bagging and Dropout $(1 + 1 + 2 = 4$ points$)$

a) Why can dropout be considered as an approximation to Bagging? Explain in two or three sentences.

b) Do we apply dropout during the inference? Justify your answer.

c) Write a pseudocode for inverted dropout implementation commenting each step.

Exercise 7.4 - Adverserial Training in NLP (2 points)

In the lecture, you were introduced to adverserial training applied in computer vision. On the internet, find how adverserial training is applied in NLP. Describe **four** adverserial techniques: give examples and short descriptions (2-3 sentences). Provide the source of your answer.

Submission instructions

The following instructions are mandatory. If you are not following them, tutors can decide to not correct your exercise.

- You have to submit the solutions of this assignment sheet as a team of 2-3 students.

- Hand in a **single** PDF file with your solutions.

- Make sure to write the student Teams ID and the name of each member of your team on your submission.

- Your assignment solution must be uploaded by only **one** of your team members to the course website.

- If you have any trouble with the submission, contact your tutor **before** the deadline.