Answer the following questions and submit a pdf file with your answers in one week.

1. How does RMSProp work for optimization with adaptive learning rates?
2. What differences do the RMSProp and Adam algorithms have?
3. What advantages does the BFGS algorithm have over Newton’s method?
4. How to apply the batch normalization method for the optimization of deep networks?
5. What’s the idea behind Polyak averaging for its application to learning problems?