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

1. Why do we apply regularization techniques on the deep networks?
2. What is weight decay, and how does it work for regularization?
3. What is the effect of L^1 norm regularization for optimization?
4. Why L^1 penalty can be used for feature selection?
5. For an object recognition task, how can we augment the data set?
6. What differences does early stopping have from other regularization methods, such as weight decay?