Alperen Karaoglu

Homework Assignment 3
Lecturer: Kyunghyun Cho
February 28, 2017

1. We have three hyptohesis sets. We start by taking the original dataset D and partition it into k equal sized blocks of data in a random manner such that each block maintains the properties of the original data set. We then pick one of the blocks of data as the validation data set and use the remaining data blocks as training data. We repeat this process k times by using each one of the data blocks once as the validation data set. We average the result from the K-folds to optain a single estimation. We go through this process for each of the three hypthosis sets and choose the best classifier by comparing their estimation. By doing this we meet our goals of determining the best classifier and its generalization performance. If the size of the original data set is not being enough such that we can't have a separate test set we can also use nested K-fold cross validation and make use of the iner data blocks to determine the best training model.

2. We have $D(y^*,M,x) = -\log p_{M*(X)} = -a_{y*}+\log \sum_{k=1}^{K} exp(a_k)$ where $a = Wx$~. We a learning rule for each column vector w_c of the weight matrix W. To do this we have to take the gradient of the Distance function with respect to the weight matrix.
$\nabla wD(y^*, M, x)$