LAB 1: SVM detector of generated domain names

Andrii Zakharchenko

November 2019

1 Assignment 1

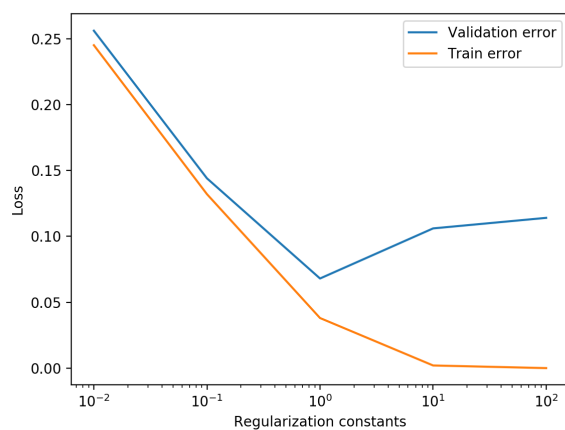| C	Train error	Validation error	Number of support vectors
0.01	0.245	0.256	1000
0.1	0.132	0.144	876
1	0.038	0.068	544
10	0.002	0.106	386
100	0.000	0.114	376

Table 1:

Figure 1: Train and validation errors with respect to regularization constants

2 Assignment 2

In table 1 we can see that the minimal validation error was achieved for regularization constant $C = 1$. The test error for corresponding SVM is $R_S(h) = 0.078$ To compute ϵ such that $R(h) \in (R_S(h) - \epsilon, R_S(h) + \epsilon)$ with propability at least $\gamma = 0.99$ we will use the following:

$$\epsilon = (L_{max} - L_{min})\sqrt{\frac{\log(2) - \log(1 - \gamma)}{2l}}$$

where $L_{max/min}$ is maximum and minimum values of loss function (which are 1 and 0 correspondingly), l is number of test examples and $\gamma = 0.99$

$$\epsilon = 0.0364$$