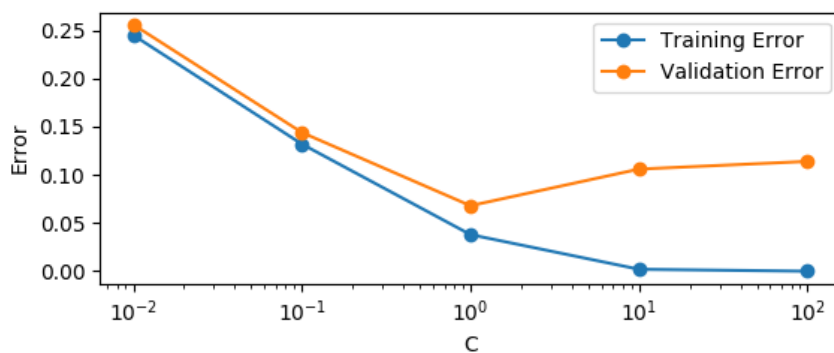# Statistical Machine Learning - SVM detector of generated domain names

Ondřej Preněk, preneond@fel.cvut.cz

10.11.2018

1 Assignment 1

Figure 1: Training errors and the validation error as a function of C in a graph.

	Training Error	Validation Error	Number of Support Vectors
C=0.01	0.245	0.256	1000.0
C=0.1	0.132	0.144	876.0
C=1.0	0.038	0.068	544.0
C=10.0	0.002	0.106	386.0
C=100.0	0.0	0.114	376.0

2 Assignment 2

The optimal C is 1.0 for which the test error is 7.85%

2.1 Confidence interval

$$R(h) \in (0.0785 - \epsilon, 0.0785 + \epsilon) \tag{1}$$

$$\epsilon = (l_{max} - lmin)\sqrt{\frac{log(2) - log(1 - \gamma)}{2l}} \tag{2}$$

$$\epsilon = (1 - 0)\sqrt{\frac{log(2) - log(1 - 0.99)}{2 * 2000}} \tag{3}$$

$$\epsilon = 0.0364 \tag{4}$$

$$R(h) \in (0.0785 - 0.0364, 0.0785 + 0.0364) \tag{5}$$

$$\tag{6}$$

3 Assignment 3

Table 1: Unnormalized SSK

	0	1	2	3
0	0.051243	0.010090	0.009503	0.009508
1	0.010090	0.067700	0.009619	0.009556
2	0.009503	0.009619	0.101678	0.017969
3	0.009508	0.009556	0.017969	0.050331

Table 2: L2-normalized SSK

	0	1	2	3
0	1.000000	0.171299	0.131650	0.187221
1	0.171299	1.000000	0.115938	0.163703
2	0.131650	0.115938	1.000000	0.251187
3	0.187221	0.163703	0.251187	1.000000