IT3105 - Exercise 1

Eirik Hammerstad & Nicklas Utgaard

November 21, 2012

Contents

1 Results

1.1 pen-digits

Classifiers	Average	St.Dev	Test Error
1 NBC	0.11	0.0	0.13
1 DTC	0.00	0.0	0.12
5 NBC	0.07	0.0	0.14
10 NBC	0.07	0.0	0.21
20 NBC	0.07	0.0	0.23
5 DTC, $depth = A$	0.0	0.0	0.13
10 DTC, $depth = 1$	0.64	0.0	0.65
10 DTC, $depth = 2$	0.309	0.003	0.31
10 DTC, $depth = A$	0.0	0.0	0.12
20 DTC, $depth = A$	0.0	0.0	0.13
5 NBC, 5 DTC	$\frac{0.138}{0.308}$	$\frac{0.0228}{0.0045}$	0.31
10 NBC, 10 DTC	$\frac{0.16}{0.30}$	$\frac{0.0323}{0.0}$	0.32
20 NBC, 20 DTC	$\frac{0.18}{0.30}$	$\frac{0.0251}{0.0}$	0.31

1.2 nursery

Classifiers	Average	St.Dev	Test Error
1 NBC	0.09	0.0	0.10
1 DTC	0.00	0.0	0.02
5 NBC	0.148	0.0363	0.17
10 NBC	0.197	0.0549	0.24
20 NBC	0.164	0.0176	0.18
5 DTC, $depth = A$	0.0	0.0	0.02
10 DTC, $depth = 1$	0.29	0.0	0.29
10 DTC, $depth = 2$	0.17	0.0	0.16
10 DTC, $depth = A$	0.00	0.0	0.02
20 DTC, $depth = A$	0.0	0.0	0.02
5 NBC, 5 DTC	$\frac{0.144}{0.17}$	$\frac{0.0321}{0.0}$	0.18
10 NBC, 10 DTC	$\frac{0.158}{0.17}$	$\frac{0.0239}{0.0}$	0.17
20 NBC, 20 DTC	$\frac{0.247}{0.17}$	$\frac{0.0715}{0.0}$	0.16

1.3 page-blocks

Classifiers	Average	St.Dev	Test Error
1 NBC	0.07	0.0	0.08
1 DTC	0.04	0.0	0.06
5 NBC	0.068	0.004	0.08
10 NBC	0.076	0.007	0.09
20 NBC	0.064	0.022	0.09
5 DTC, $depth = A$	0.05	0.0	0.05
10 DTC, $depth = 1$	0.08	0.0	0.09
10 DTC, $depth = 2$	0.07	0.0	0.08
10 DTC, $depth = A$	0.04	0.0	0.06
20 DTC, $depth = A$	0.04	0.0	0.07
5 NBC, 5 DTC	$\frac{0.066}{0.04}$	$\frac{0.0167}{0.0}$	0.06
10 NBC, 10 DTC	$\frac{0.075}{0.07}$	$\frac{0.007}{0.0}$	0.08
20 NBC, 20 DTC	$\frac{0.0765}{0.08}$	$\frac{0.006}{0.0}$	0.08

1.4 glass

Classifiers	Average	St.Dev	Test Error
1 NBC	0.2	0.0	0.53
1 DTC	0.07	0.0	0.44
5 NBC	0.302	0.0715	0.47
10 NBC	0.261	0.0223	0.58
20 NBC	0.289	0.0223	0.65
5 DTC, $depth = A$	0.09	0.0	0.47
10 DTC, $depth = 1$	0.50	0.0	0.60
10 DTC, $depth = 2$	0.291	0.003	0.51
10 DTC, $depth = A$	0.06	0.0	0.35
20 DTC, $depth = A$	0.08	0.0	0.44
5 NBC, 5 DTC	$\frac{0.23}{0.418}$	$\frac{0.0255}{0.0716}$	0.60
10 NBC, 10 DTC	$\frac{0.308}{0.445}$	$\frac{0.0316}{0.0474}$	0.42
20 NBC, 20 DTC	$\frac{0.325}{0.32}$	$\frac{0.0390}{0.0}$	0.56

1.5 yeast

Classifiers	**Average**	**St.Dev**	**Test Error**
1 NBC	0.36	0.0	0.42
1 DTC	0.19	0.0	0.52
5 NBC	0.424	0.0391	0.49
10 NBC	0.423	0.0279	0.53
20 NBC	0.483	0.0492	0.48
5 DTC, $depth = A$	0.18	0.0	0.54
10 DTC, $depth = 1$	0.58	0.0	0.36
10 DTC, $depth = 2$	0.49	0.0	0.54
10 DTC, $depth = A$	0.17	0.0	0.58
20 DTC, $depth = A$	0.19	0.0	0.53
5 NBC, 5 DTC	$\frac{0.428}{0.51}$	$\frac{0.0676}{0.0}$	0.53
10 NBC, 10 DTC	$\frac{0.437}{0.49}$	$\frac{0.0352}{0.0}$	0.51
20 NBC, 20 DTC	$\frac{0.472}{0.48}$	$\frac{0.042}{0.0}$	0.52

2 Analysis

We will here go through the dataset and point out some of the interesting trends that occurs.

2.1 Overfitting

In section 1.1 on page 3 looking at the *pen-digits* dataset we see an interesting trend in that it seems like one *NBC* classifier works better then twenty classifier. This also occurs in nursery and yeast. A reason for this might be overfitting of the later classifiers, which might occur when our learner starts to memorize trainingdata instead of generalizing from trend [1].
As cited by Alexander Vezhnevets and Olga Barinove, "... overfitting is induced by fitting so called 'confusing samples', that are samples misclassified by 'perfect' Bayesian classifier. Overfitting in boosting seems to occur only when target distributions overlap or the noise is present ..."[2]

2.2 Error propagation

A thing we saw clearly when running our tests was that the error ϵ gradually became smaller and smaller for each classifier, which in turn increased the weights Ω_C of each classifier, making the first classifiers insignificant compared the latter ones.

Generally speaking we saw a big problem with DTC and boosting, where the changes in the instance weights ω_i was in almost all cases so small that the entropy S stayed pretty much the same, resulting equivalent classifiers. This was especially the case on 'easy' datasets, where the DTC classifier had a very low error on the training data.

[1] Wikipedia, Overfitting_(Machine_learning)

[2] `http://www.inf.ethz.ch/personal/vezhneva/Pubs/AvoidingBoostingOverfitting.pdf`