CS156b Project1

Letter Recognition — Test

Ling Li
Learning Systems Group, Caltech
December 4, 2001

Final test

Three "candidates" were produced by AdaBoost.M2 with 16-70-50-26 neural nets on the first 12000 train samples, and were "validated" on the rest 3997 samples.

They used different training algorithms for individual neural nets. Due to different learning speeds and time, parameters used were different:

method	epoch	# nets	time (h)	ν (%)	π_t (%)
SG	1000	100	66.5	2.752	2.348
CG	600	80	113	2.777	2.298
MM	1000	100	n/a	3.252	2.873

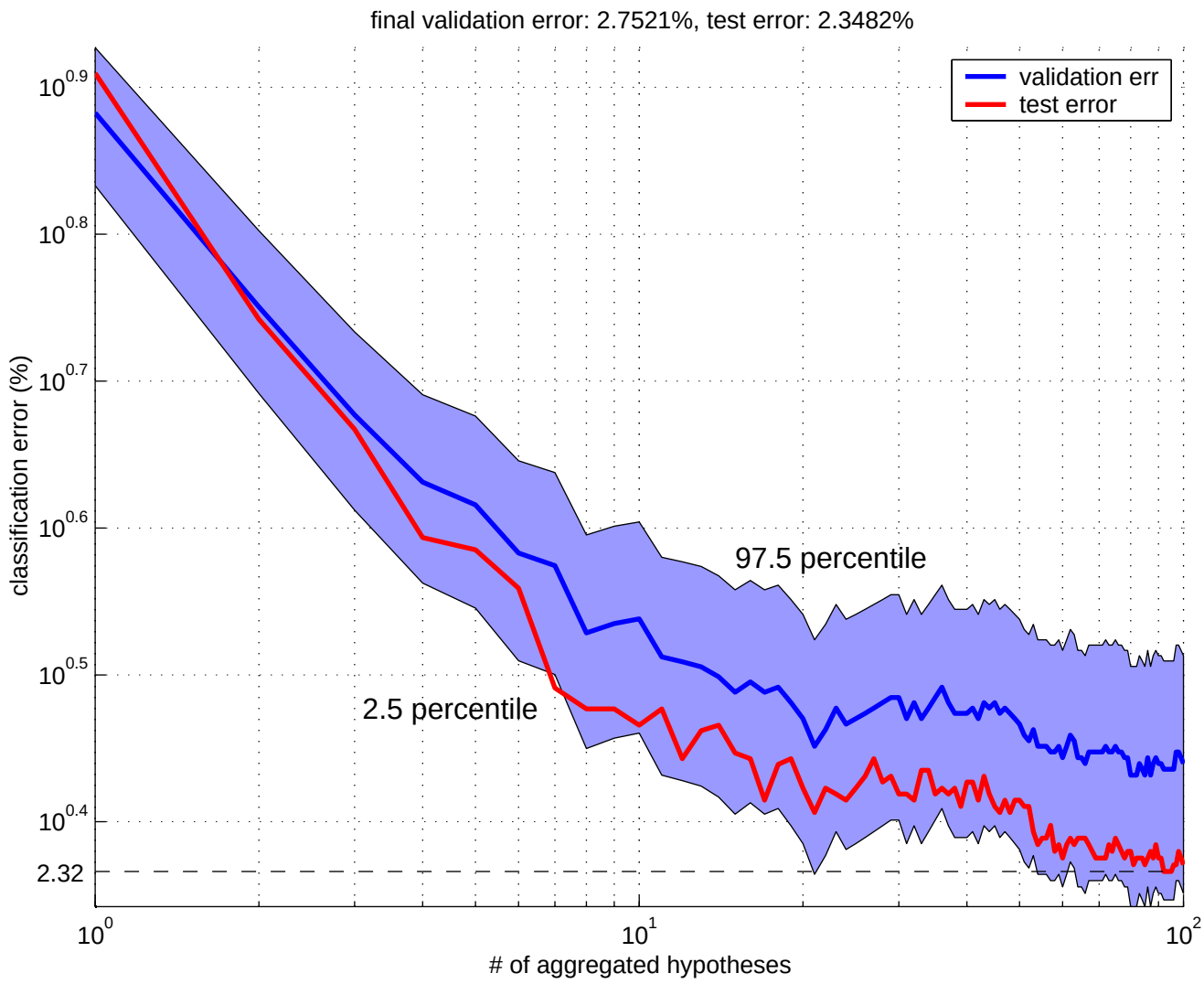
The one with the best "validation error" (using SG) was selected. I showed the 95% confidence interval (in the last presentation) is $[2.24\%, 3.26\%]$.

However, from the test results, we will see:

- the one with the best test error is CG,
  (note that π_t is also an estimate of π. It is not π.)

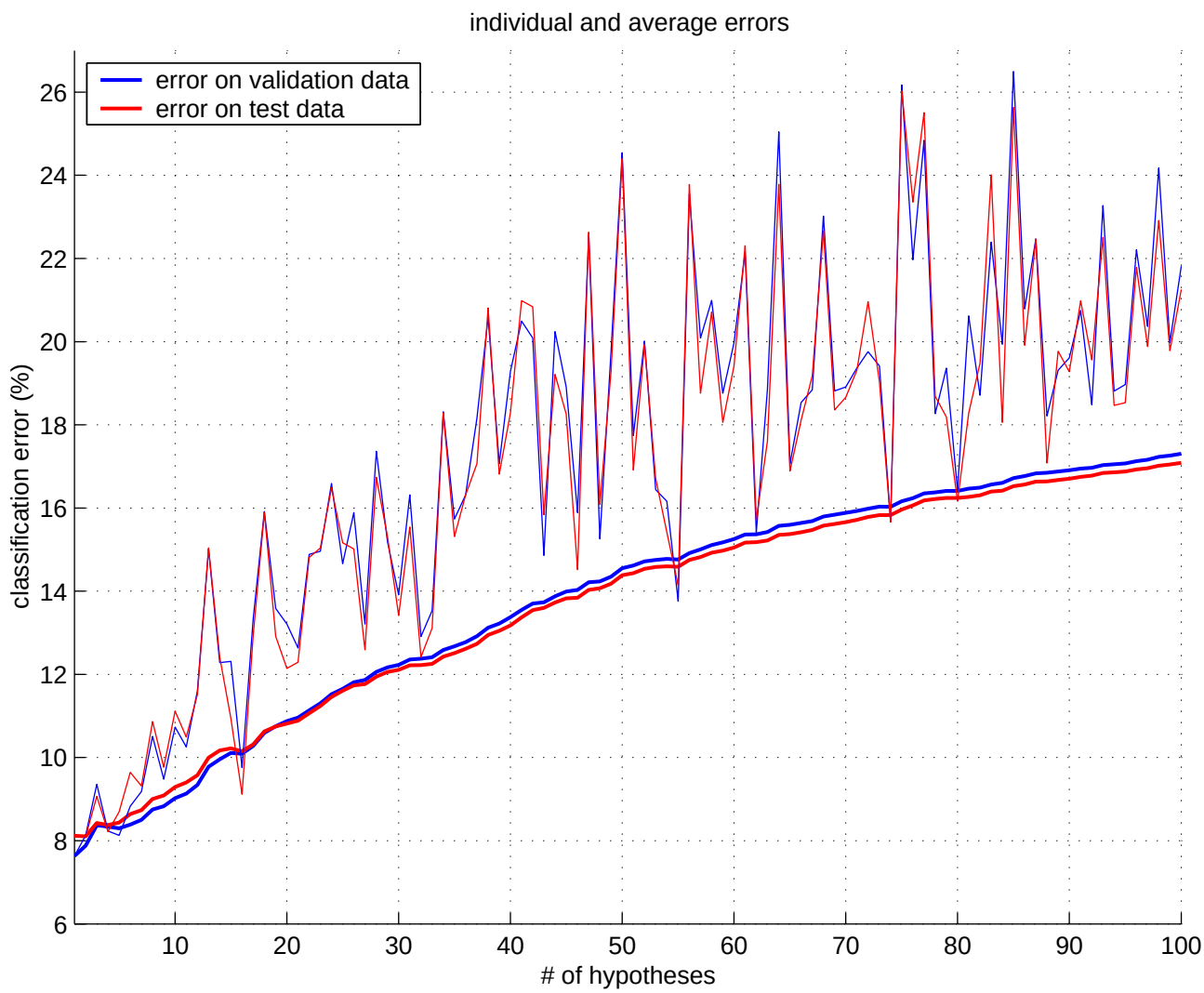
- the "worst one" (MM) has the best generalization.

AdaBoost.M2 with SG

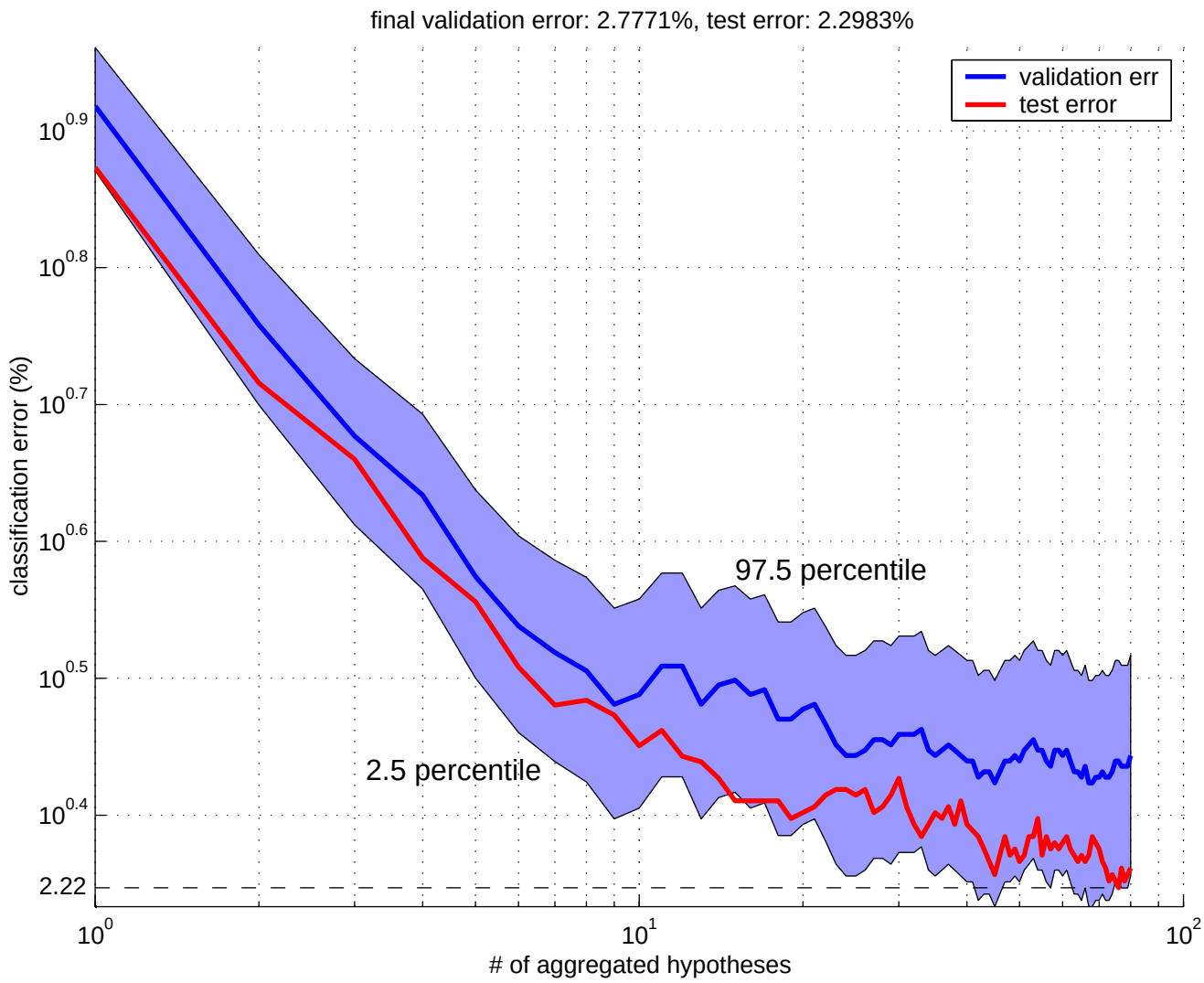


Curious about the generalization of individual nets?

I was also curious. Below I plotted the errors of individual nets on the validation / test sets.



AdaBoost.M2 with CG



AdaBoost.M2 with MM

It seems that this has the best generalization...