CS57300: Homework 3

Long Zhen, lzhen@purdue.edu

March 9, 2015

	isFunny				*isPositive*			
	zero − one − loss		*baseline*		*zero − one − loss*		*baseline*	
sample%	*mean*	*std*	*mean*	*std*	*mean*	*std*	*mean*	*std*
10%	0.3669	0.0054	0.510	0.002	0.2236	0.0304	0.508	0.016
50%	0.3457	0.0122	0.504	0.006	0.1896	0.0091	0.512	0.005
90%	0.3353	0.0262	0.502	0.011	0.1807	0.0165	0.513	0.014

Table 1: Q3(b) Table for avg. and std. of zero-one-loss across all sampling and their baseline default error stats

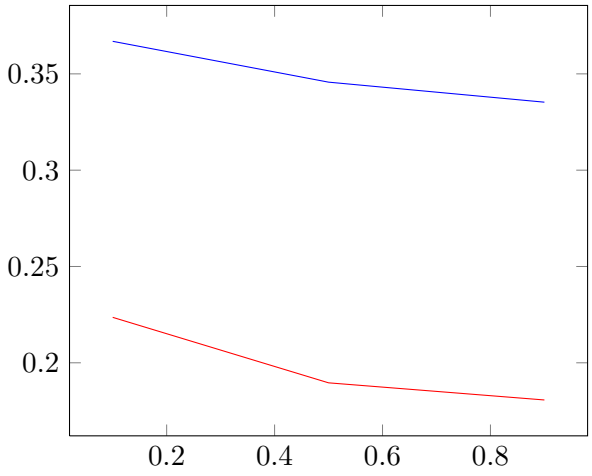

Figure 1: training data size vs. zero-one-loss, blue: *isFunny*, red: *isPositive*

Discussion

- As we can see, with the size of training data increasing, we can have better score (less error) to predict the test data.
- Just like what we learned from the lecture, NBC have very good prediction accuracy (checking both means) and is very stable (checking std).