

Naive Bayes case study

- Training set: 10,000 emails that are either SPAM or HAM
- Testing set: 1,000 additional emails
- Train a Naive Bayes classifier on (a subset of) the training set
- Predict SPAM/HAM on the test set and compute accuracy.

Mark – I am working with the East power desk to purchase space for an EnronOnline banner ad on a PJM website. We are buying 7 ads at 500/month/ad for 3 months (\$10,500 total). They are running this ad as a pilot program offered for only 3 months. I am attaching the agreement they sent to us. I would like to revise section 2.01 to state that EnronOnline has first right of refusal to keep the ad on their site if they extend the program after three months. Could you help me revise this agreement?

Thanks

Kal

Mark – I am working with the East power desk to purchase space for an EnronOnline banner ad on a PJM website. We are buying 7 ads at 500/month/ad for 3 months (\$10,500 total). They are running this ad as a pilot program offered for only 3 months. I am attaching the agreement they sent to us. I would like to revise section 2.01 to state that EnronOnline has first right of refusal to keep the ad on their site if they extend the program after three months. Could you help me revise this agreement?

Thanks

Kal

HAM!

Body Wrap at Home to lose 6-20 inches in one hour. With Bodywrap we guarantee: you'll lose 6-8 Inches in one hour 100% Satisfaction or your money back. Bodywrap is soothing formula that contours, cleanses and rejuvenates your body while reducing inches.

ambuscade eunice diffeomorphism sycamore kampala excelled possessor
dobbin aqueduct tertiary smudgy beebread shawnee flat anybody multi
necromancy harriet seder amherst paleozoic jejune irredentism cornet
buckley eleanor casteth ponce administrate babysitter admittance
abernathy bethesda busy joaquin casebook unidimensional carboloy
captious bracelet anniversary edwin albumin tangent

Body Wrap at Home to lose 6-20 inches in one hour. With Bodywrap we guarantee: you'll lose 6-8 Inches in one hour 100% Satisfaction or your money back;BRçj/Pç Bodywrap is soothing formula that contours, cleanses and rejuvenates your body while reducing inches.jBRç ambuscade eunice diffeomorphism sycamore kampala excelled possessor dobbin aqueduct tertiary smudgy beebread shawnee flat anybody multi necromancy harriet seder amherst paleozoic jejune irredentism cornet buckley eleanor casteth ponce administrate babysitter admittance abernathy bethesda busy joaquin casebook unidimensional carboloy captious bracelet anniversary edwin albumin tangent

SPAM!

Non-trivial HAM words

enron	8.58508e+00
scott	6.50723e+00
chris	6.43892e+00
edison	6.13924e+00
jeff	6.10057e+00
disclosure	5.97333e+00
mw	5.94861e+00
pge	5.92610e+00
karen	5.89284e+00
kimberly	5.82908e+00

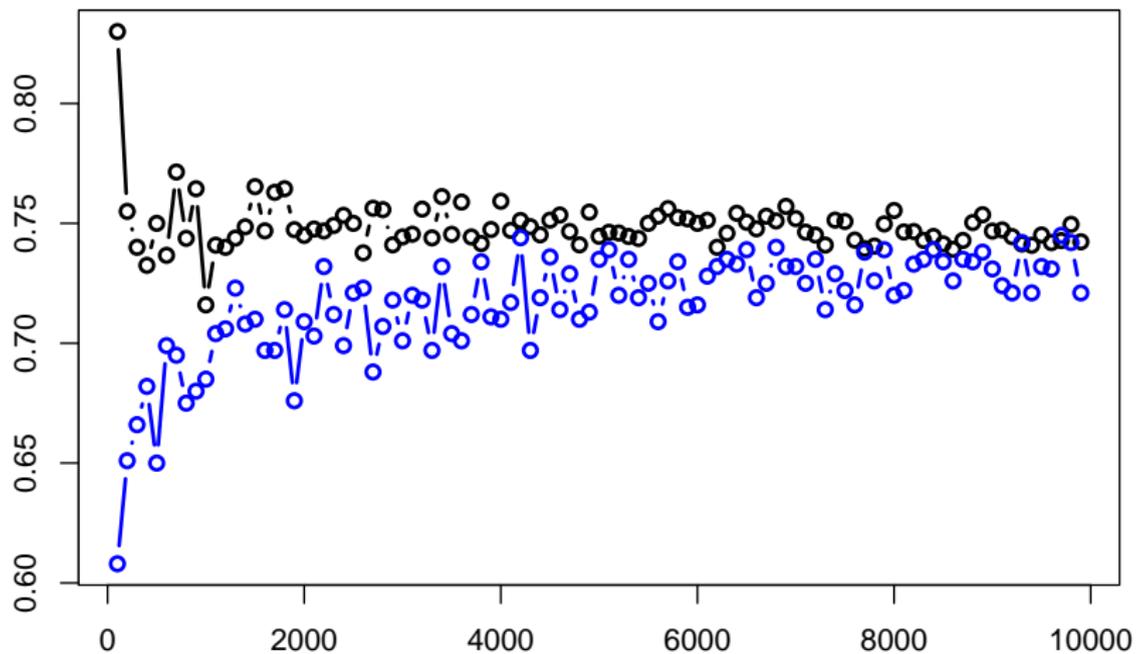
Non-trivial SPAM words

taacaeecorpenroncom	8.14474e+00
ur	7.80475e+00
contentdtexthtml	7.50449e+00
multipart	7.11542e+00
nds	7.10469e+00
ger	7.10006e+00
thr	7.10006e+00
reas	7.09384e+00
bgcolordffffff	7.05898e+00
tdtd	7.01361e+00

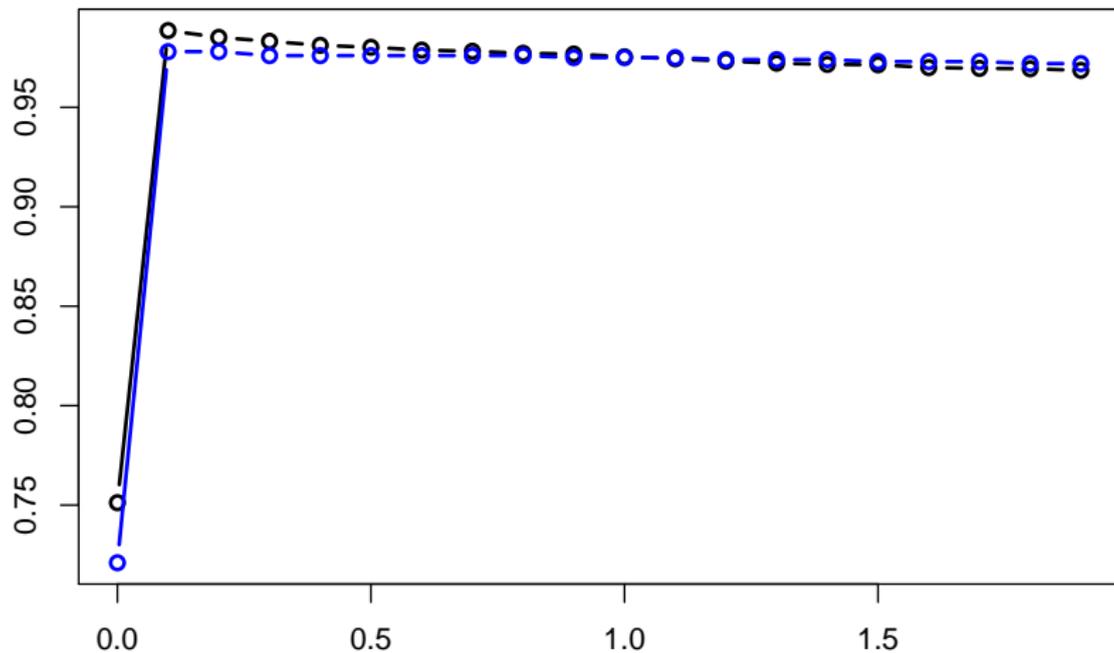
More non-trivial SPAM words

bilion	6.51536e+00
namedgenerator	6.44339e+00
tras	6.40845e+00
illustrator	6.36260e+00
contentdmshtml	6.20141e+00
meds	6.18801e+00
wastes	6.15868e+00
omit	6.14268e+00
pills	6.02968e+00
spe	5.99834e+00
mime	5.99445e+00

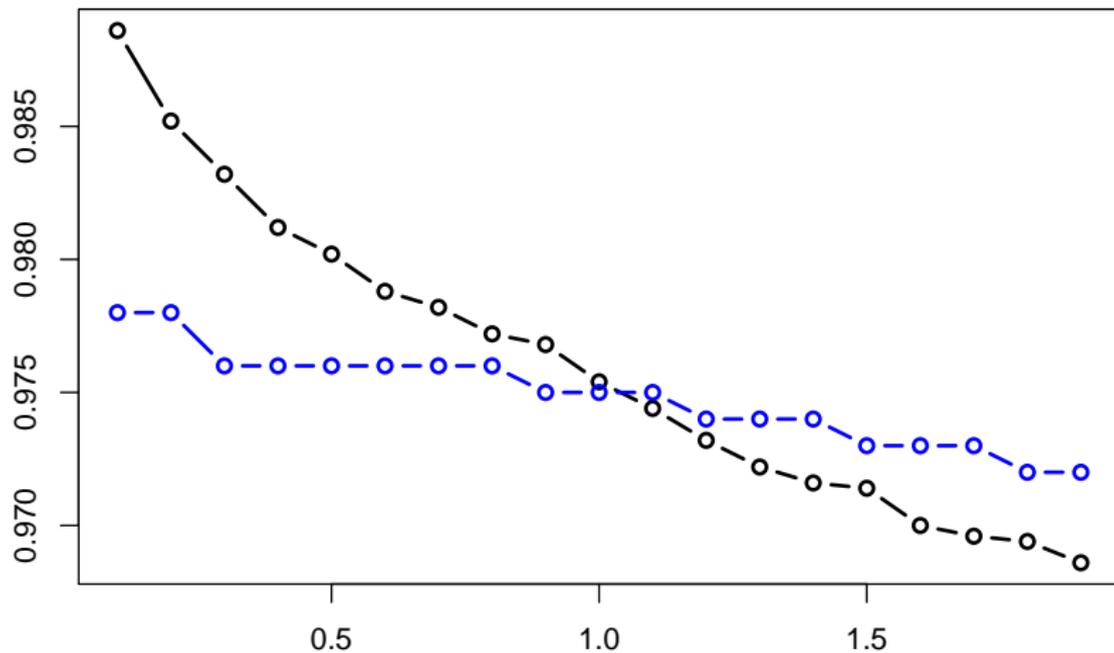
Sensitivity to training size



Sensitivity to smoothing



Sensitivity to smoothing



SPAM words (0.1 smoothing)

gbbl	1.05488e+01
widthd	9.83269e+00
heightd	9.64469e+00
borderd	9.40989e+00
geec	9.02820e+00
cellpaddingd	8.96986e+00
voip	8.87144e+00
cellspacingd	8.86078e+00
hotfix	8.77111e+00
ur	8.60916e+00

HAM words (0.1 smoothing)

ferc	7.82131e+00
enrons	7.60930e+00
scott	7.45650e+00
pipeline	7.33990e+00
chris	7.29062e+00
enron	7.18227e+00
ena	7.13472e+00
joe	7.07833e+00
yards	6.96004e+00