

1. Choose a blog or a newsfeed (or something similar with an Atom or RSS feed). It should be on a topic or topics of which you are qualified to provide classification training data. Find something with at least 100 entries.

Create between four and eight different categories for the entries in the feed:
examples:

work, class, family, news, deals
liberal, conservative, moderate, libertarian
sports, local, financial, national, international, entertainment
metal, electronic, ambient, folk, hip-hop, pop

Download and process the pages of the feed as per the week 12 class slides.

1A. Had difficulty finding a rss feed as I am not an avid rss/atom reader. The best feed I could find was the Reddit Subreddit XML feeds that have plenty of content.

I chose to utilize the subreddit /r/funny. I have been an avid reddit and imgur user for very long. I've noticed a pattern in the content so much so that a majority of it can be predicted from the title and categorized:

- meme - any sort of captioned text with an iconic picture.
- gif - animated content, can be real or edited content
- animal - anything related to animals - can span media categories
- original - undoctored pictures taken of content excluding animals
- edited - doctored content by photo editing software excluding animals
- media - content from tv, magazine, media, movies
- monty_python - it's a noteworthy category

2. Manually classify the first 50 entries, and then classify (using the fisher classifier) the remaining 50 entries. Report the `cprob()` values for the 50 titles as well. From the title or entry itself, specify the 1-, 2-, or 3-gram that you used for the string to classify. Do not repeat strings; you will have 50 unique strings. For example, in these titles the string used is marked with *s:

Create a table with the title, the string used for classification, `cprob()`, predicted category, and actual category.

2a. A table was created utilizing manually classification. Given this title, hopefully a trained model would be able to predict the content of the web. Commonly people utilize similar tags, titles, and nomenclatures when sharing content with a broad dataset in a short frame of time.

Here is Data[0], it is recognized as an animal gif of a sloth hilariously eating. I manually train the model that the title is indicative of an animal picture. The following was preformed for data 0-49. The subsequent 50-99 was then predicted by the model.
Training the data:

```
cl.train(lines[0], 'animal')
```

Fig 1a. Data 0



Fig 1b. Data 50



Title: "Hungry... But opening mouth is too much work."

Link:

http://www.reddit.com/r/funny/comments/1squt0/hungry_but_opening_mouth_is_too_much_work

The model had difficulty discerning from the initial training pattern good options. The model trained itself that the majority of pictures would be 'original' or 'animal' pictures which was generally not the case. Wishfully with the title "Macaulay Culkin... woof!", the model should classify it as animal. Given more data this may become the case.

See Data

See figure 2: for results

3. Assess the performance of your classifier in each of your categories by computing precision and recall. Note that the definitions are slightly different in the context of classification; see:

http://en.wikipedia.org/wiki/Precision_and_recall#Definition_.28classification_context.29

Here is the analysis of the information from figure 2.

For classification tasks, the terms **true positives**, **true negatives**, **false positives**, and **false negatives**.

Laymans terms and localizing it to this problem.

- **true positives** - Predicted the correct classification matching to actual **prediction**
- **true negatives** - Predicted the correct - incorrect classification for correctly classified content
- **false positives** - the prediction was correct but the actual content was classified incorrectly
- **false negatives** - the prediction was incorrect but the actual content was classified incorrectly

Prediction accuracy is the accuracy based off its prediction of that category vs the actual

The confusion matrix provided above, in my opinion true positives are provided, and true negatives would be the inverse.

False positives and false negatives are difficult to determine in this scheme without other human classifiers participating to determine if incorrect classifications were made.

Figure 3

Category	predicted	actual	True Positive	Prediction Accuracy
meme	4	9	1/4	25.00%
gif	0	6	6/0	0.00%
animal	27	3	2/27	7.41%
original	17	16	06/17	35.29%
edited	1	8	0/0	0.00%
media	1	7	1/0	0.00%
monty_python	0	1	0/0	0.00%

Figure 2

num	title	cprob ()	predicted category	actual category	links
50	Macaulay Culkin... woof!	0	animal	gif	http://www.reddit.com/r/funny/comments/1spgyw/macaulay_culkin_woof/
51	How would you not know?!	0	animal	media	http://www.reddit.com/r/funny/comments/1spqnt/how_would_you_not_know/
52	Does this make anyone else happy? [GRUMPY CAT]	0	animal	meme	http://www.reddit.com/r/funny/comments/1sr2h2/does_this_make_anyone_else_happy_grumpy_cat/
53	They see me bowling	0	animal	gif	http://www.reddit.com/r/funny/comments/1spp1n/they_see_me_bowling/
54	Grammar Matters	0	original	edited	http://www.reddit.com/r/funny/comments/1sr0pv/grammar_matters/
55	We did it. We made the yahoo homepage again.	0	animal	media	http://www.reddit.com/r/funny/comments/1sptrg/we_did_it_we_made_the_yahoo_homepage_again/
56	Pretty much sums it up	0	animal	media	http://www.reddit.com/r/funny/comments/1srfje/pretty_much_sums_it_up/
57	Superb...!	0	animal	animal	http://www.reddit.com/r/funny/comments/1sprer/superb/
58	30,000 Feet	0	original	edited	http://www.reddit.com/r/funny/comments/1spt5x/30000_feet/
59	This kid is going places	0	edited	original	http://www.reddit.com/r/funny/comments/1spymp/this_kid_is_going_places/
60	That guy with the iPhone cookies in his car seems to have gotten a ticket	0	original	original	http://www.reddit.com/r/funny/comments/1sq5t7/that_guy_with_the_iphone_cookies_in_his_car_seems_to_have_gotten_a_ticket/
61	A girl posted this to instagram titled 'Breakfast Bear'. I didn't have the heart to tell her who it resembled.	0	original	original	http://www.reddit.com/r/funny/comments/1sqfp6/a_girl_posted_this_to_instagram_titled_breakfast_bear_i_didnt_have_the_heart_to_tell_her_who_it_resembled/
62	They sure are...	0	original	meme	http://www.reddit.com/r/funny/comments/1spoja/they_sure_are/
63	That awkward moment when you	0	animal	original	http://www.reddit.com/r/funny/comments/1sq54w/that_awkward_moment_when_you_take_fashion_advice/

Stanley Zheng
December 12th 2013
Assignment 10

	take fashion advice from the public trains				
64	Death's list	0	original	edited	http://www.reddit.com/r/funny/comments/1sq75y/deaths_list/
65	Got this in the mail.	0	animal	meme	http://www.reddit.com/r/funny/comments/1sqpmv/got_this_in_the_mail/
66	Magic	0	original	meme	http://www.reddit.com/r/funny/comments/1sps3c/magic/
67	Can I get you anything?	0	animal	meme	http://www.reddit.com/r/funny/comments/1splsk/can_i_get_you_anything/
68	Seems legit	0	animal	original	http://www.reddit.com/r/funny/comments/1sq2vp/seems_legit/
69	lets talk about your future	0	animal	meme	http://www.reddit.com/r/funny/comments/1sq9p9/lets_talk_about_your_future/
70	Santacat and I	0	original	meme	http://www.reddit.com/r/funny/comments/1sqmyx/santacat_and_i/
71	What's the ONE thing you want more than anything?	0	meme	edited	http://www.reddit.com/r/funny/comments/1sqobe/whats_the_one_thing_you_want_more_than_anything/
72	First day of school, 1984.	0	animal	original	http://www.reddit.com/r/funny/comments/1sphp5/first_day_of_school_1984/
73	fuuu	0	original	media	http://www.reddit.com/r/funny/comments/1sppdf/□/
74	The Messiah's car	0	animal	edited	http://www.reddit.com/r/funny/comments/1spohb/the_messiahs_car/
75	Hey girl, feel my sweater	0	media	meme	http://www.reddit.com/r/funny/comments/1spzdi/hey_girl_feel_my_sweater/
76	Nice Day For A Cruise	0	animal	original	http://www.reddit.com/r/funny/comments/1spp1e/nice_day_for_a_cruise/
77	Ooooh Snap!	0	animal	gif	http://www.reddit.com/r/funny/comments/1spi6m/ooooh_snap/
78	Classic Futurama	0	original	media	http://www.reddit.com/r/funny/comments/1sphwt/classic_futurama/
79	Eric, what did I say about calling your sister the Devil?	0	original	media	http://www.reddit.com/r/funny/comments/1spr4i/eric_what_did_i_say_about_calling_your_sister_the_devil/
80	New coworker has this on her desk. She know's what's up!	0	animal	original	http://www.reddit.com/r/funny/comments/1sq70h/new_coworker_has_this_on_her_desk_she_knows_whats/
81	And I will love it and squeeze it and hug it.	0	animal	original	http://www.reddit.com/r/funny/comments/1spvlq/and_i_will_love_it_and_squeeze_it_and_hug_it/
82	Why women make men productive [fixed]	0	meme	meme	http://www.reddit.com/r/funny/comments/1sr2b1/why_women_make_men_productive_fixed/
83	This Geese	0	original	gif	http://www.reddit.com/r/funny/comments/1sql6i/this_geese_tsunami_is_a_shit_storm/

Stanley Zheng
December 12th 2013
Assignment 10

	Tsunami is a shit storm				
84	I was using a calculator app on my iPhone when I remembered the "8008135" Upside-Down thing people use to do in school. I tried it and this happened. Mind Blown	0	animal	edited	http://www.reddit.com/r/funny/comments/1sql3i/i was using a calculator app on my iphone when i/
85	Grumpy cat-doge	0	animal	edited	http://www.reddit.com/r/funny/comments/1sq6qt/grumpy_catdoge/
86	Can you cook?	0	animal	gif	http://www.reddit.com/r/funny/comments/1spt0a/can_you_cook/
87	Dog Sneeze	0	animal	animal	http://www.reddit.com/r/funny/comments/1spv0a/dog_sneeze/
88	Mad Cat!	0	original	animal	http://www.reddit.com/r/funny/comments/1sq8xf/mad_cat/
89	Seriously, I can't be the only one seeing this.	0	animal	media	http://www.reddit.com/r/funny/comments/1sqndr/seriously i cant be the only one seeing this/
90	Every Lesbian Porno Ever	0	original	gif	http://www.reddit.com/r/funny/comments/1sqp43/every_lesbian_porno_ever/
91	When my friend went to the 22 concert in Nashville fans kept coming up to her thinking she was actually Taylor Swift	0	original	original	http://www.reddit.com/r/funny/comments/1spgp7/when my friend went to the 22 concert in/
92	Follow Up To The iPhone Cookies	0	original	original	http://www.reddit.com/r/funny/comments/1srgf6/follow up to the iphone cookies/
93	I am not sure if she has read the back of the shirt	0	meme	original	http://www.reddit.com/r/funny/comments/1sra7q/i am not sure if she has read the back of t he/
94	This Just happened faith in humanity restored!!	0	original	original	http://www.reddit.com/r/funny/comments/1sq3re/this just happened faith in humanity restored/
95	I was quite photogenic	0	meme	original	http://www.reddit.com/r/funny/comments/1srd8w/i was quite photogenic as a child/

Stanley Zheng
December 12th 2013
Assignment 10

	as a child.				
96	This preschool lets their kids wrestle.. just follow a few rules.	0	original	original	http://www.reddit.com/r/funny/comments/1sqk55/this_preschool_lets_their_kids_wrestle_just/
97	I found these in my grandpa's house, can't really figure it out.	0	animal	original	http://www.reddit.com/r/funny/comments/1spjlf/i_found_these_in_my_grandpas_house_cant_reall/
98	Apparently my math teacher's having women problems.	0	edited	real	http://www.reddit.com/r/funny/comments/1srall/apparently_my_math_teachers_having_women_problems/
99	Time to walk funny.	0	animal	monty_python	http://www.reddit.com/r/funny/comments/1sps3v/time_to_walk_funny/
100	Wall off to an early lead...	0	animal	original	http://www.reddit.com/r/funny/comments/1sq4po/wall_off_to_an_early_lead/