Stanley Zheng
December 12th 2013
Assignment 10

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

Stanley Zheng
December 12th 2013
Assignment 10

2. Manually classify the first 50 entries, and then classify (usingthe 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 toclassify. 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

Stanley Zheng
December 12th 2013
Assignment 10

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 correctly
- **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%

Stanley Zheng
December 12th 2013
Assignment 10

Figure 2

num	title	cprob()	predicted category	actual category
50	Macaulay Culkin... woof!	0	animal	gif
51	How would you not know?!	0	animal	media
52	Does this make anyone else happy? [GRUMPY CAT]	0	animal	meme
53	They see me bowling	0	animal	gif
54	Grammar Matters	0	original	edited
55	We did it. We made the yahoo homepage again.	0	animal	media
56	Pretty much sums it up	0	animal	media
57	Superb...!	0	animal	animal
58	30,000 Feet	0	original	edited
59	This kid is going places	0	edited	original
60	That guy with the iPhone cookies in his car seems to have gotten a ticket	0	original	original
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
62	They sure are...	0	original	meme
63	That awkward moment when you take fashion advice from the public trains	0	animal	original
64	Death's list	0	original	edited
65	Got this in the mail.	0	animal	meme
66	Magic	0	original	meme
67	Can I get you anything?	0	animal	meme
68	Seems legit	0	animal	original
69	lets talk about your future	0	animal	meme
70	Santacat and I	0	original	meme
71	What's the ONE thing you want more than anything?	0	meme	edited
72	First day of school, 1984.	0	animal	original
73	fuuu	0	original	media
74	The Messiah's car	0	animal	edited
75	Hey girl, feel my sweater	0	media	meme
76	Nice Day For A Cruise	0	animal	original
77	Ooooh Snap!	0	animal	gif
78	Classic Futurama	0	original	media
79	Eric, what did I say about calling your	0	original	media

	sister the Devil?			
80	New coworker has this on her desk. She know's what's up!	0	animal	original
81	And I will love it and squeeze it and hug it.	0	animal	original
82	Why women make men productive [fixed]	0	meme	meme
83	This Geese Tsunami is a shit storm	0	original	gif
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
85	Grumpy cat-doge	0	animal	edited
86	Can you cook?	0	animal	gif
87	Dog Sneeze	0	animal	animal
88	Mad Cat!	0	original	animal
89	Seriously, I can't be the only one seeing this.	0	animal	media
90	Every Lesbian Porno Ever	0	original	gif
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
92	Follow Up To The iPhone Cookies	0	original	original
93	I am not sure if she has read the back of the shirt	0	meme	original
94	This Just happened faith in humanity restored!!	0	original	original
95	I was quite photogenic as a child.	0	meme	original
96	This preschool lets their kids wrestle.. just follow a few rules.	0	original	original
97	I found these in my grandpa's house, can't really figure it out.	0	animal	original
98	Apparently my math teacher's having women problems.	0	edited	real
99	Time to walk funny.	0	animal	monty_python
100	Wall off to an early lead...	0	animal	original

Stanley Zheng
December 12th 2013
Assignment 10