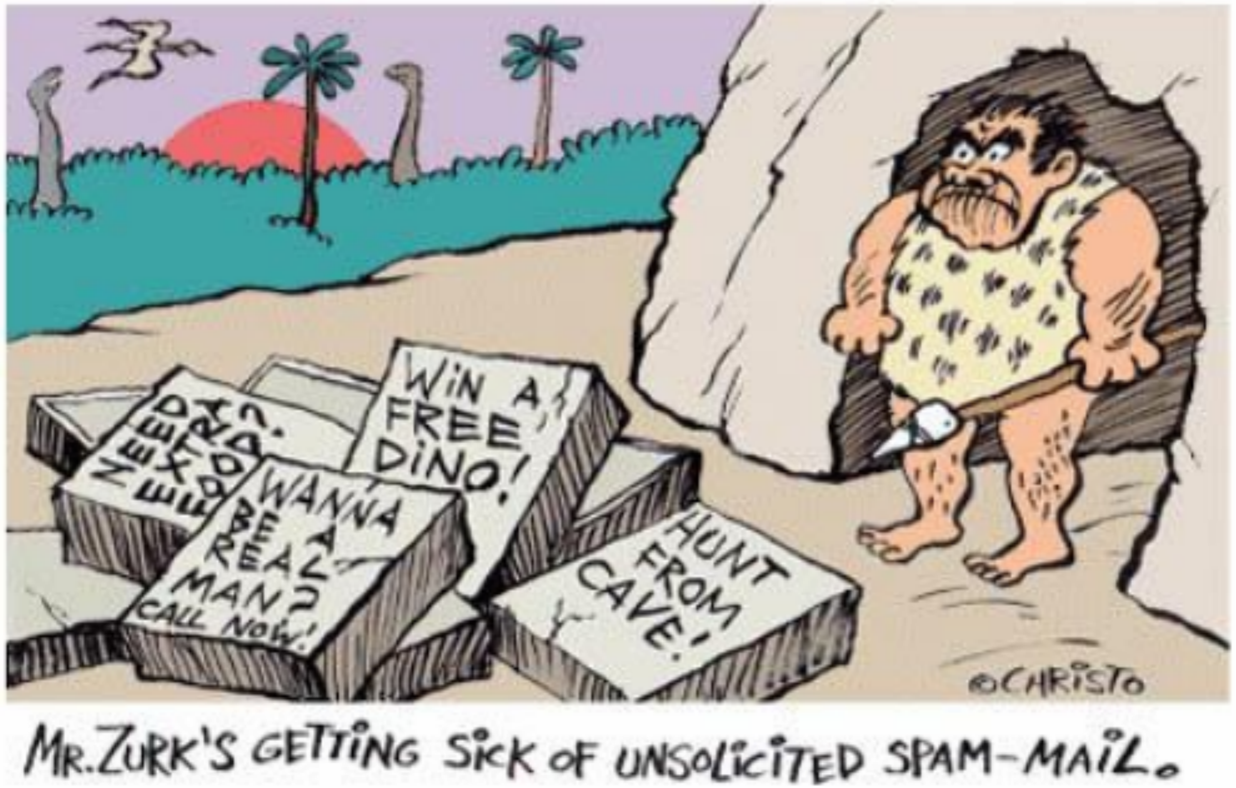
1 101100001000 101100001000 101100001000 101100001000 101100001000 1

CSCI 2824: Discrete Structures

Lecture 29: Applications of Bayes Theorem

Rachel Cox

Department of Computer Science


NEED DATA? FOR EXTRA RED
WIN A FREE DINO!
WANNA BE A REAL MAN? CALL NOW!
HUNT FROM CAVE!
©CHRISTO
MR. ZURK'S GETTING SICK OF UNSOLICITED SPAM-MAIL.

0 101100001000 101100001000 101100001000 101100001000 101100001000 0

Applications of Bayes' Theorem

Let E and F be events from sample space S such that $p(E) \neq 0$ and $p(F) \neq 0$. Then we have:

$$p(F|E) = \frac{p(E|F)p(F)}{p(E)} = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\bar{F})p(\bar{F})}$$

Bayesian Reasoning – updating our prior belief about an event $\left(p(F)\right)$ based on an observation or new knowledge of another event E.

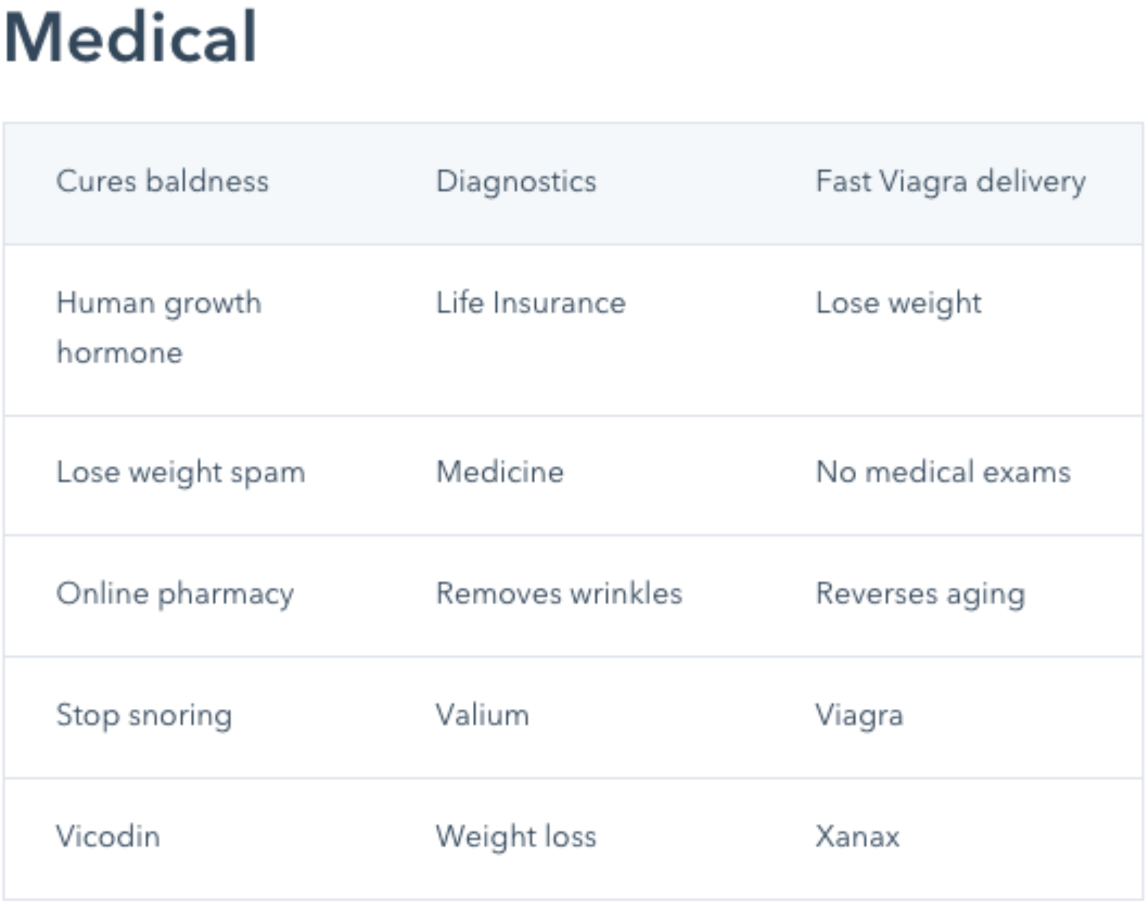
One possible application is an email SPAM filter.

- The SPAM filter updates our belief of the probability that an email is SPAM or not based on the event that particular words show up in the email.

Applications of Bayes' Theorem – SPAM filters

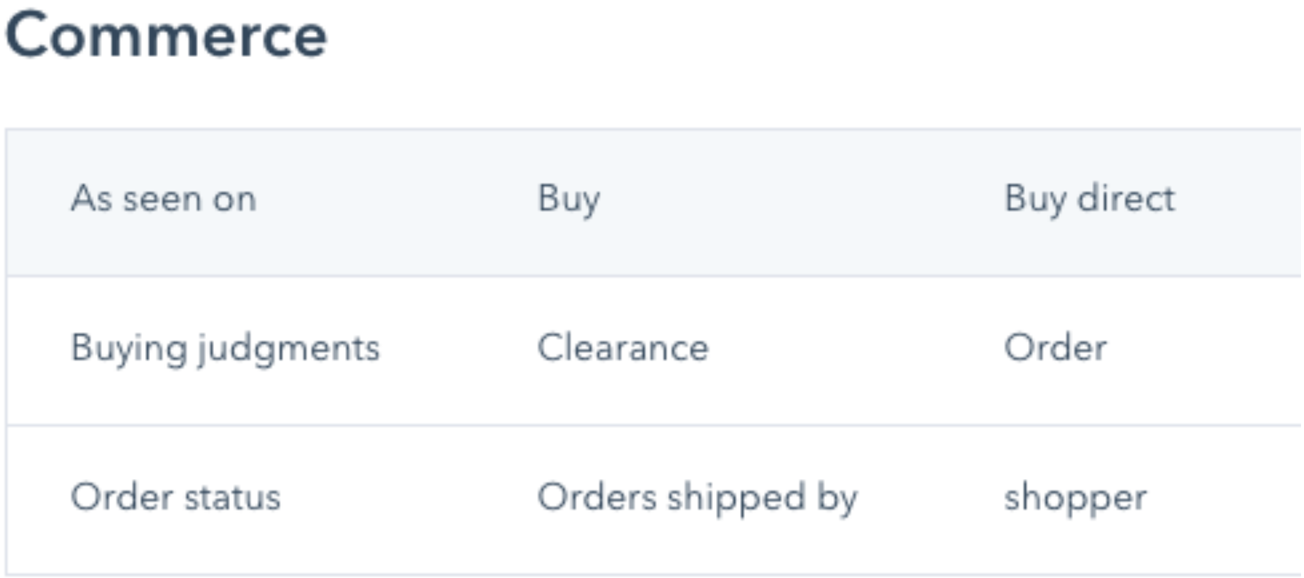
Bayesian spam filters (BSF) use information about previously seen emails to predict whether or not an incoming email is SPAM.

Simple BSFs might look for occurrences of particular words in the message that would indicate the message is spam.



Medical

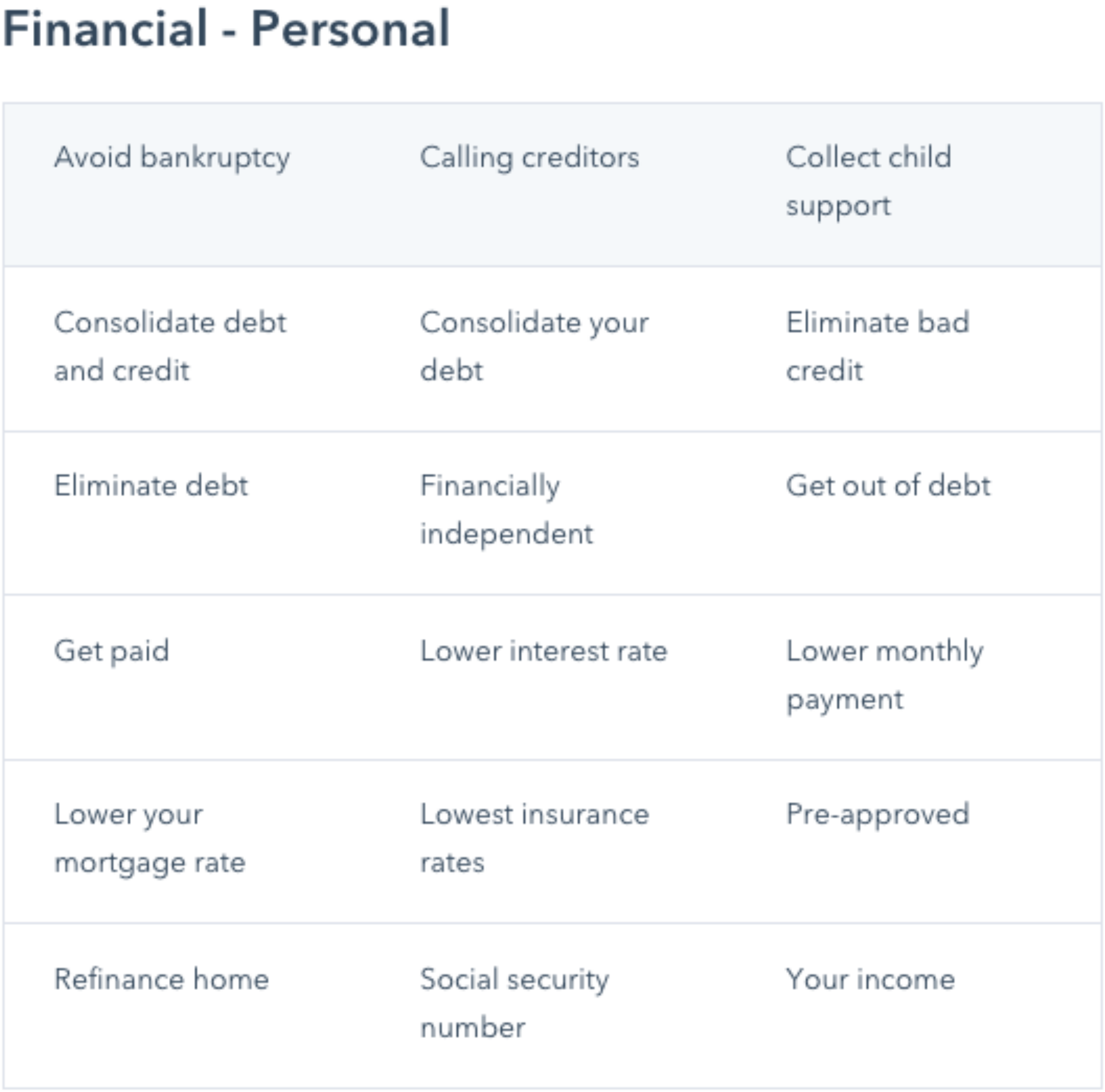
Cures baldness	Diagnostics	Fast Viagra delivery
Human growth hormone	Life Insurance	Lose weight
Lose weight spam	Medicine	No medical exams
Online pharmacy	Removes wrinkles	Reverses aging
Stop snoring	Valium	Viagra
Vicodin	Weight loss	Xanax



Commerce

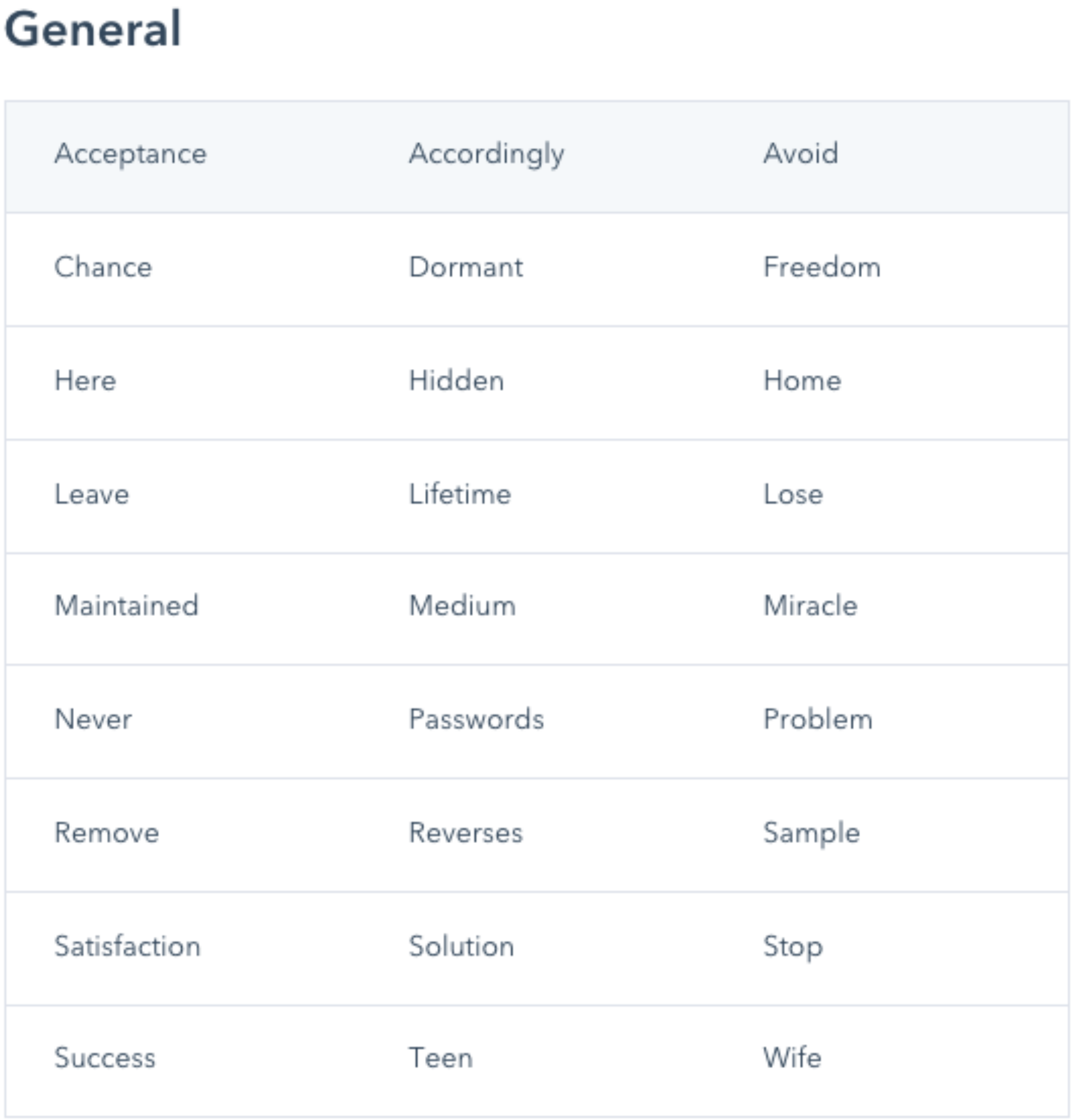
As seen on	Buy	Buy direct
Buying judgments	Clearance	Order
Order status	Orders shipped by	shopper

Applications of Bayes’ Theorem – SPAM filters



Financial - Personal

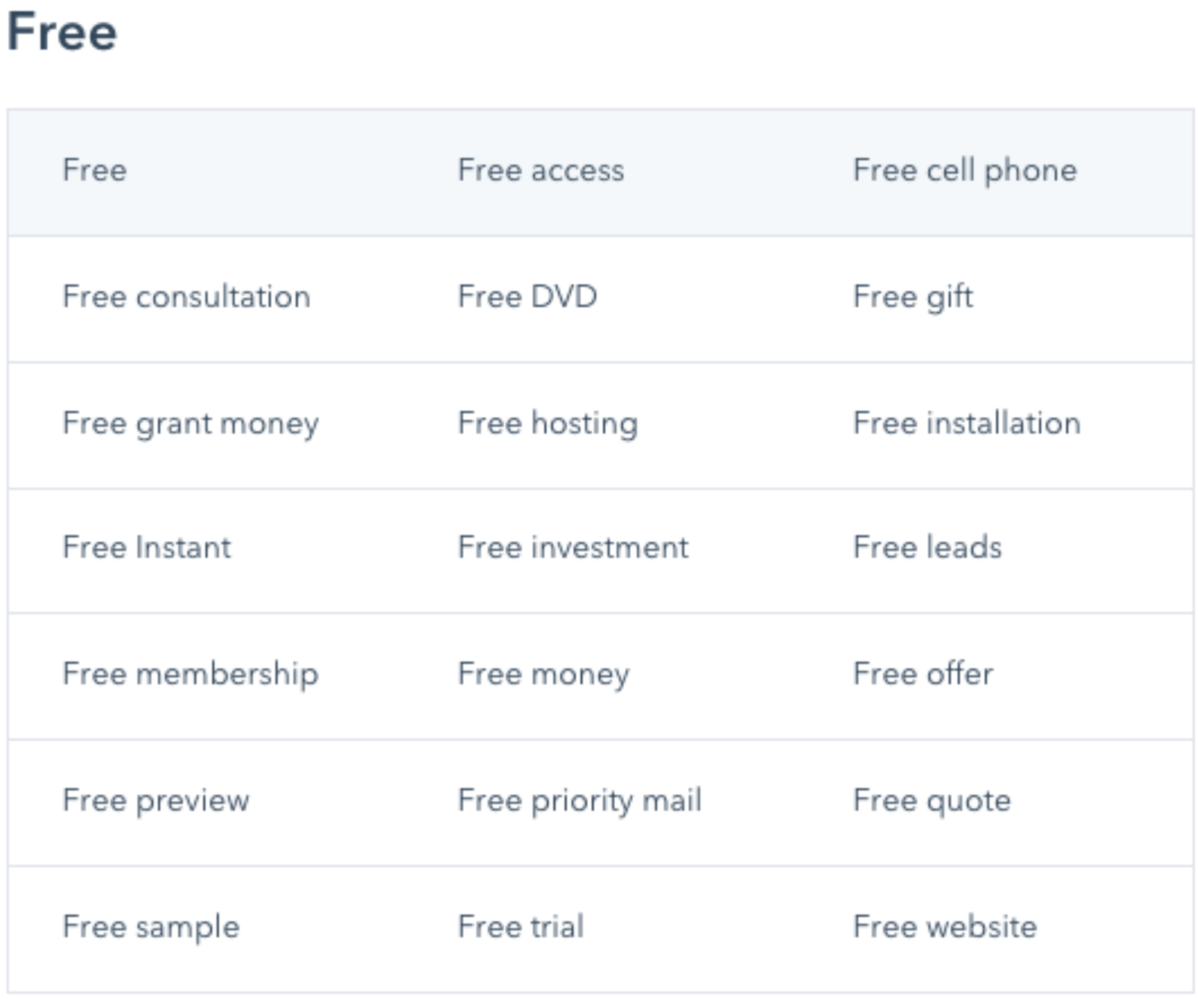
Avoid bankruptcy	Calling creditors	Collect child support
Consolidate debt and credit	Consolidate your debt	Eliminate bad credit
Eliminate debt	Financially independent	Get out of debt
Get paid	Lower interest rate	Lower monthly payment
Lower your mortgage rate	Lowest insurance rates	Pre-approved
Refinance home	Social security number	Your income



General

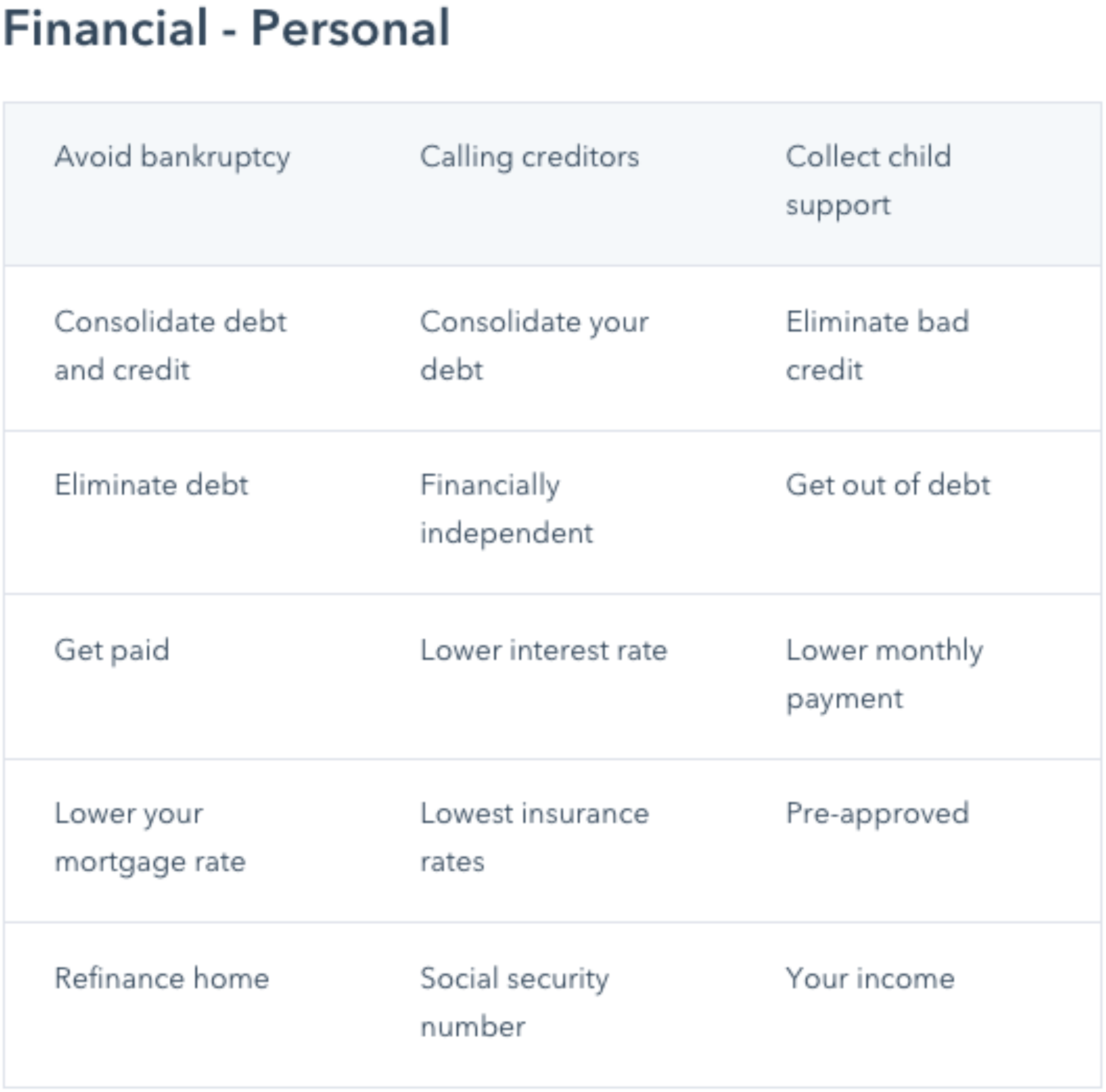
Acceptance	Accordingly	Avoid
Chance	Dormant	Freedom
Here	Hidden	Home
Leave	Lifetime	Lose
Maintained	Medium	Miracle
Never	Passwords	Problem
Remove	Reverses	Sample
Satisfaction	Solution	Stop
Success	Teen	Wife

Applications of Bayes' Theorem – SPAM filters



Free

Free	Free access	Free cell phone
Free consultation	Free DVD	Free gift
Free grant money	Free hosting	Free installation
Free Instant	Free investment	Free leads
Free membership	Free money	Free offer
Free preview	Free priority mail	Free quote
Free sample	Free trial	Free website



Financial - Personal

Avoid bankruptcy	Calling creditors	Collect child support
Consolidate debt and credit	Consolidate your debt	Eliminate bad credit
Eliminate debt	Financially independent	Get out of debt
Get paid	Lower interest rate	Lower monthly payment
Lower your mortgage rate	Lowest insurance rates	Pre-approved
Refinance home	Social security number	Your income

Applications of Bayes' Theorem – SPAM filters

Suppose you've catalogued a bunch of emails. Some of them as SPAM, and some as actual emails that you'd want to read.

Could you estimate the probability that the email is SPAM, given that the word w appears in an email?

Want $\boldsymbol{\underbrace{p(SPAM \mid w\,)}}$

"It's not the most sophisticated Spam blocker I've tried, but it's the only one that works!"

Applications of Bayes' Theorem – SPAM filters

$$\underbrace{p(spam\,|w)} = \frac{\overset{\downarrow}{p(w\,|spam)}\;\overset{\downarrow}{p(spam)}}{\underset{\uparrow}{p(w)}} = \frac{p(w\mid spam)\,p(spam)}{p(w\,|spam)\,p(spam) + p(w\mid ham)\,p(ham)}$$

Note: We're letting $\overset{\downarrow}{\overline{spam}} = ham$

Applications of Bayes' Theorem – SPAM filters

$$p(w \mid spam) = \frac{|W|}{|S|} = \frac{\text{\# spam messages containing } w}{\text{total \# of spam messages}}$$

$$p(w \mid ham) = \frac{|W|}{|H|} = \frac{\text{\# ham messages containing } w}{\text{total \# of ham messages}}$$

What about $p(spam)$ and $p(ham)$?

Applications of Bayes’ Theorem – SPAM filters

$$p(spam) = \frac{\text{\# spam messages}}{\text{total \# messages}}$$

$$p(ham) = \frac{\text{\# ham messages}}{\text{total \# messages}}$$

Applications of Bayes' Theorem – SPAM filters

Example: Given the following set of messages, estimate $p(spam\,|fly)$ and $p(ham\,|fly)$

spam	spam	spam	ham	ham
nigeria	fly	money	money	mom
fly	buy	buy	buy	fly
money	nigeria	pills	work	nigeria

Law of Total Prob.

Bayes' Theorem

$$p(spam\,|fly\,) = \frac{p\,(\,fly\,|spam\,)\,p\,(spam)}{p\,(fly)} = \frac{p\,(fly\,|\,spam\,)\,p\,(spam\,)}{p\,(\,fly\,|spam\,)\,p\,(spam\,) + p\,(fly\,|ham\,)\,p\,(ham\,)}$$

$$p(ham\,|fly\,) = \frac{p\,(\,fly\,|ham\,)\,p\,(ham)}{p\,(fly)} = \frac{p\,(fly\,|\,ham\,)\,p\,(ham\,)}{p\,(\,fly\,|spam\,)\,p\,(spam\,) + p\,(fly\,|ham\,)\,p\,(ham\,)}$$

Applications of Bayes' Theorem – SPAM filters

Example: Given the following set of messages, estimate $p(spam\ |fly)$ and $p(ham\ |fly)$

spam	spam	spam	ham	ham
nigeria	fly	money	money	mom
fly	buy	buy	buy	fly
money	nigeria	pills	work	nigeria

$$p(spam\ |fly\ ) = \frac{p\ (fly\ |\ spam\ )\ p\ (spam\ )}{p\ (\ fly\ |spam\ )\ p\ (spam\ ) + p\ (fly\ |ham\ )\ p\ (ham\ )} = \frac{\frac{2}{3}\cdot\frac{3}{5}}{\frac{2}{3}\cdot\frac{3}{5}+\frac{1}{2}\cdot\frac{2}{5}} = \frac{\frac{2}{5}}{\frac{2}{5}+\frac{1}{5}} = \frac{2}{3}$$

$$p(ham\ |fly\ ) = \frac{p\ (fly\ |\ ham\ )\ p\ (ham\ )}{p\ (\ fly\ |spam\ )\ p\ (spam\ ) + p\ (fly\ |ham\ )\ p\ (ham\ )} = \frac{\frac{1}{2}\cdot\frac{2}{5}}{\frac{2}{3}\cdot\frac{3}{5}+\frac{1}{2}\cdot\frac{2}{5}} = \frac{\frac{1}{5}}{\frac{3}{5}} = \frac{1}{3}$$

Applications of Bayes' Theorem – SPAM filters

How should we classify an email that we receive, based only on the word fly?

Since $p(\, spam \,|fly\,) > 0.5$, it's reasonable to classify the email as spam, but we can also set a threshold different from 0.5 based on how bad it would be to make a classification mistake.

- Classifying spam as not spam, is annoying.
- Classifying an important message as spam could be terrible!

In practice, maybe we'd set the threshold for SPAM classification higher than 0.5.

Applications of Bayes' Theorem – SPAM filters

We just classified an email based on one word. In practice, we may want to classify an email with as many words as possible simultaneously.

We need to make a key assumption about the relationship of words in emails: Words are ***conditionally independent***, given their category.

Events A and B are ***conditionally independent*** given Y, if and only if

$$p(A \cap B \mid Y) = p(A \mid Y) \cdot p(B \mid Y)$$

independence : $p(A \cap B) = p(A) \cdot p(B)$

e.g. $p(fly \textbf{ and } nigeria \mid spam) = \underbrace{p(fly \mid spam)}_{} \cdot \underbrace{p(nigeria \mid spam)}_{}$

Applications of Bayes' Theorem – SPAM filters

Example: $p(fly \textbf{ and } nigeria \mid spam) = p(fly|spam) \cdot p(nigeria|spam)$

spam	**spam**	**spam**	**ham**	**ham**
nigeria	fly	money	money	mom
fly	buy	buy	buy	fly
money	nigeria	pills	work	nigeria

We could estimate $p(spam \mid \{fly, nigeria\})$ and $p(ham \mid \{fly, nigeria\})$ by counting up the instances of both "fly" and "nigeria", in spam or ham emails.

However:

1. We don't want to have to re-estimate all the specific combinations of words whenever we get a new email.
2. We might not have many emails with both words.

Applications of Bayes' Theorem – SPAM filters

Note: The denominators are the same for $p(\ ham\ |email\ )$ and $p(spam\ |email\ )$:

$$p\ (\ ham\ |email\ ) = \frac{p\ (\ email\ |ham\ )\ p\ (ham\ )}{p\ (email\ )}$$

$$p\ (\ spam\ |email\ ) = \frac{p\ (\ email\ |\ spam\ )\ p\ (spam\ )}{p\ (email\ )}$$

- ❖ We can classify an email entirely based on which numerator is larger.

Applications of Bayes’ Theorem – SPAM filters

e.g. $\frac{1}{5} < \frac{4}{5}$

Example: Based on this example, predict whether the email containing “fly” and “nigeria” is spam or ham.

spam	spam	spam	ham	ham
nigeria	fly	money	money	mom
fly	buy	buy	buy	fly
money	nigeria	pills	work	nigeria

use conditional independence

$$p(spam\,|fly \cap nigeria\,) = \frac{p\,(fly \cap nigeria\,|spam\,)\,p\,(spam\,)}{(denominator)} = \frac{p(fly|spam)\cdot p(nigeria|spam)\,p(spam)}{(denominator)}$$

$$= \frac{2}{3}\cdot\frac{2}{3}\cdot\frac{3}{5}\ \boxed{=\frac{4}{15}}$$

$$p(ham\,|fly \cap nigeria\,) = \frac{p\,(fly \cap nigeria\,|ham\,)\,p\,(ham\,)}{(denominator)} =$$

$$= p(fly|ham)\cdot p(nigeria|ham)\,p(ham)$$

$$= \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{2}{3}\ \boxed{=\frac{1}{10}}$$

We used Bayesian Reasoning to update our belief about the probability of an email being ham or spam.

- Start with our prior beliefs. $p(spam)$ or $p(ham)$
- Make observations of particular words in an incoming email, and want to update our belief about spam or ham based on these observations.
  $p(spam\ |'words\ in\ email')$ and $p(ham\ |'words\ in\ email')$
- Bayes’ Theorem allows us to calculate these based on the words in old emails that we know are spam or ham.

Next: Recursion!