Prioritizing What to Work On

Building a Spam Classifier
Advice for Applying Machine Learning:

From: cheapsales@buystufffromme.com

To: ang@cs.stanford.edu

Subject: Buy now!

Deal of the week! Buy now! Rolex w4tchs - \$100 Med1cine (any kind) - \$50 Also low cost M0rgages available.

Span

From: Alfred Ng

To: ang@cs.stanford.edu

Subject: Christmas dates?

Hey Andrew,
Was talking to Mom about plans
for Xmas. When do you get off
work. Meet Dec 22?
Alf

Non-spok

From: cheapsales@buystufffromme.com

To: ang@cs.stanford.edu

Subject: Buy now!

Deal of the week! Buy now!
Rolex w4tchs - \$100
Medicine (any kind) - \$50
Also low cost M0rgages
available.

Span

From: Alfred Ng

To: ang@cs.stanford.edu

Subject: Christmas dates?

Hey Andrew,
Was talking to Mom about plans
for Xmas. When do you get off
work. Meet Dec 22?
Alf

Non-spam

From: cheapsales@buystufffromme.com

To: ang@cs.stanford.edu

Subject: Buy now!

Deal of the week! Buy now!
Rolex w4tchs - \$100
Medicine (any kind) - \$50
Also low cost M0rgages
available.

Spam (1)

From: Alfred Ng

To: ang@cs.stanford.edu

Subject: Christmas dates?

Hey Andrew,
Was talking to Mom about plans
for Xmas. When do you get off
work. Meet Dec 22?
Alf

Non-spom (0)

Supervised learning. x = features of email. y = spam (1) or not spam (0).

Features x: Choose 100 words indicative of spam/not spam.

Supervised learning. $x = \text{features of email.} \ y = \text{spam (1) or not spam (0)}.$

Features x: Choose 100 words indicative of spam/not spam.

E.g. deal, buy, discont, andrew, now, ...

Supervised learning. $x = \text{features of email.} \ y = \text{spam (1) or not spam (0)}.$

Features x: Choose 100 words indicative of spam/not spam.

E.g. deal, buy, discont, andrew, now, ...

From: cheapsales@buystufffromme.com

To: ang@cs.stanford.edu

Subject: Buy now!

Deal of the week! Buy now!

Windows'u Etkinleştir Windows'u etkinleştirmek için Ayarlar'a gidin.

Andrew Ng

Supervised learning. $x = \text{features of email.} \ y = \text{spam (1) or not spam (0)}.$

Features x: Choose 100 words indicative of spam/not spam.

E.g. deal, buy, discont, andrew, now, ...

andrew buy deal discont now

From: cheapsales@buystufffromme.com

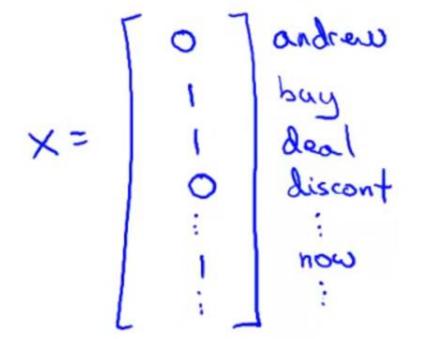
To: ang@cs.stanford.edu

Subject: Buy now!

Deal of the week! Buy now!

Supervised learning. $x = \text{features of email.} \ y = \text{spam (1) or not spam (0)}.$

Features x: Choose 100 words indicative of spam/not spam.





From: cheapsales@buystufffromme.com

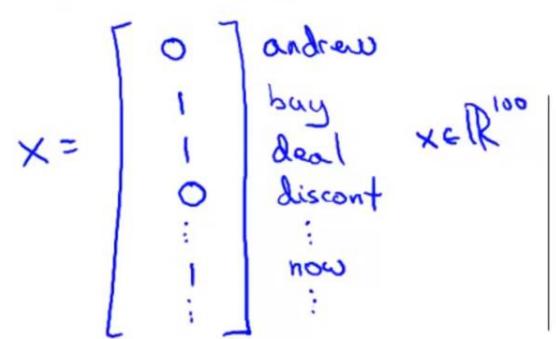
To: ang@cs.stanford.edu

Subject: Buy now!

Deal of the week! Buy now!

Supervised learning. $x = \text{features of email.} \ y = \text{spam (1) or not spam (0)}.$

Features x: Choose 100 words indicative of spam/not spam.



From: cheapsales@buystufffromme.com

To: ang@cs.stanford.edu

Subject: Buy now!

Deal of the week! Buy now!

Supervised learning. x = features of email. y = spam (1) or not spam (0).

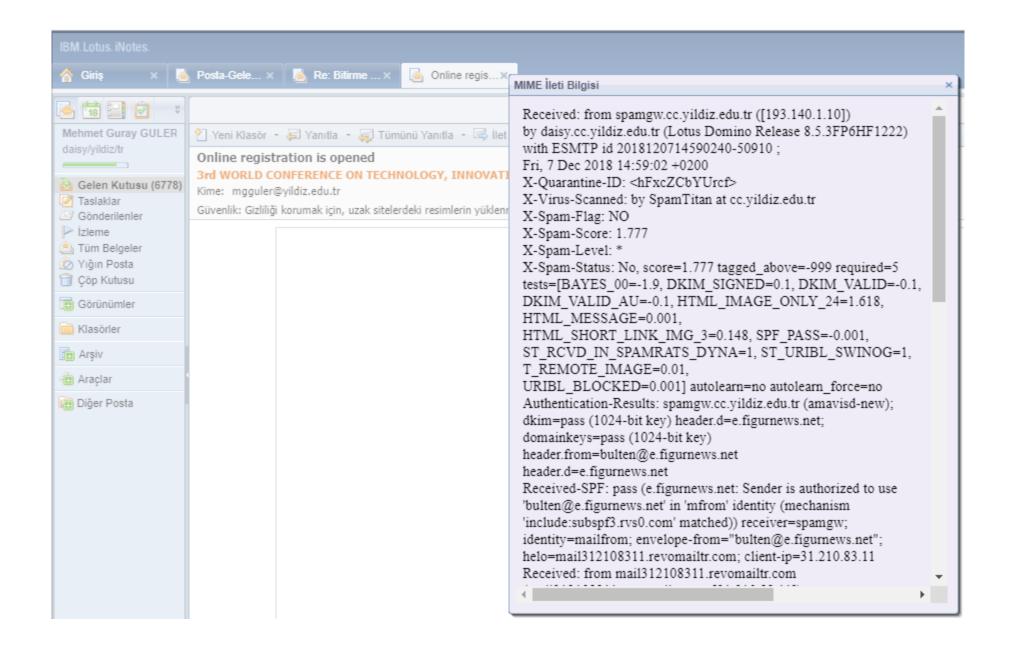
Note: In practice, take most frequently occurring n words (10,000 to 50,000) in training set, rather than manually pick 100 words.

How to spend your time to make it have low error?

- Collect lots of data
 - E.g. "honeypot" project.

How to spend your time to make it have low error?

- Collect lots of data
 - E.g. "honeypot" project.
- Develop sophisticated features based on email routing information (from email header).



How to spend your time to make it have low error?

- Collect lots of data
 - E.g. "honeypot" project.
- Develop sophisticated features based on email routing information (from email header).
- Develop sophisticated features for message body, e.g. should "discount" and "discounts" be treated as the same word? How about "deal" and "Dealer"? Features about punctuation?



How to spend your time to make it have low error?

- Collect lots of data
 - E.g. "honeypot" project.
- Develop sophisticated features based on email routing information (from email header).
- Develop sophisticated features for message body, e.g. should "discount" and "discounts" be treated as the same word? How about "deal" and "Dealer"? Features about punctuation?
- Develop sophisticated algorithm to detect misspellings (e.g. m0rtgage, med1cine, w4tches.)

