

Gaussian kernel  $f_i = \text{sim}(x^{(i)}, l^{(j)}) = \exp\left(-\frac{\|x^{(i)} - l^{(j)}\|^2}{2\sigma^2}\right)$

1.2.1 gaussiankernel.m

sim = gaussiankernel(x1, x2, sigma)

$z = x1 - x2;$

$z = z^T \cdot z;$

sim = exp(-z / (2 \* sigma \* sigma))

1.2.3 Example Dataset 3

$c$  と  $\sigma$  を 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30 に変更しそれぞれ試す

(8種類)<sup>2</sup>

for 変数 = [ 配列 ]

...

end for

if 条件

end if

式1 0 = 式2  
式1 ~ = 式2 }  $\rightarrow$   $\sigma$  が小さいと式1と式2が等しくないとき1, 等しいとき0

## 2. Spam classification

機械学習のタスクを始める前に、データセット中の例を見ておくことは役に立つ。

2.1 processEmail.m は、前処理 (preprocessing) を行っている

str ... string (single word)

vocabList ... 配列 (stringの)

strが vocabList 中に存在したそのindexを word-indices に追加する。  
しなれば、skip

word\_indices = processEmail(email\_contents)

1次元配列の長さ length(1次元配列)

strcmp(文字列1, 文字列2) → 1 or 0 (等しい時1)

for文を途中で抜けるのは break

## 2.2 Extracting Features from Email

vocabList = [ str<sup>1</sup> str<sup>2</sup> ... str<sup>n</sup> ]

x-n中に存在し、存在する

x = [ 0 1 0 ... 1 0 ]

x = emailFeatures(word\_indices)

Week 7 Ex 6 3/3

2.3 spamTrain.m ... 4000 training examples  
spamTest.m ... 1000 test examples

各 Email は processEmail , emailFeatures を用いて  $x^{(i)} \in \mathbb{R}^{1899}$  <sup>n</sup> <sub>1899</sub> に変換した