Spam or Ham (Bayes Naive klassifikation)

Ved dette projekt ses på hvordan man kan lave et simpelt spamfilter. Så ud fra oplysninger om en mail skal det automatisk afgøres om det er spam og derved ikke vises i mail-box, eller om det er Ham(non-spam).

Kom selv i gang med at læse dokumentet Naive-Bayes.pdf og forstå ideerne og principperne bag Bayes Naive klassifikation, og undervejs skal du selv arbejde med følgende datasæt, hvor der er set på 35000 mails som man har forskellige info om (oprindelse, mailadresse, indhold) og for hver har bestemt om det er spam eller ham:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Oprindelse |  |  |  | Mail |  |  |  | Indhold |  |
|  | DK | Europa(-DK) | Usa | Andet | Firma | Google | Hotmail mm. | Andet | dating | Gevinster og spil | Andet |
| spam | 20% | 30% | 35% | 55% | 10% | 20% | 60% | 80% | 80% | 90% | 12,5% |
| Ham | 80% | 70% | 65% | 45% | 90% | 80% | 40% | 20% | 20% | 10% | 87,5% |
| antal | 10000 | 12000 | 8000 | 5000 | 17000 | 6450 | 5400 | 6150 | 4325 | 4975 | 25700 |

Opgave 1:

a)Hvis man blot modtager en mail fra en Hotmail-konto vil man da tænke det er spam eller ham ?

b)Hvis man blot modtager en tilfældig mail vil man da tænke det er spam eller ham?

c) Forklar hvorfor det kan være svært at afgøre om det er spam hvis det er mail fra DK fra en firma-mail og hvor indhold er relateret til dating.

2: Indfør selv stoksatiske variable X1,X2, X3 og Y og angiv udfaldsrum for hver af dem.

3: Opstil posterior forholdet og forklar det smarte der er sket via bayes formel. Redegør ligeledes for betydningen af forholdet og hvorfor man ser på om det er over eller under 1.

4: Angiv antagelsen som anvendes ved Bayes Naive klassifikation og forklar hvad det gør for beregning af sandsynligheder.

5: Bestem prior forholdet og beregn vægten w0

6: bestem alle betingede sandsynligheder for hver enkelt information givet at det er hhv. spam og ham. (i alt 22 sandsynligheder) og forklar ideen bag en af disse udregninger.

7: For hver af de 11 informationer bestemmes forholdet mellem sandsynlighederne

8: forklar hvordan man kommer fra resultaterne fra 7 til vægte og udregn vægte hørende til hver af informationerne (i alt 11).

9: Forklar hvad der sker ved at benytte logaritme på udtrykket hvor man ellers ganger faktorer sammen.

10: afgør hvilke informationer der taler for spam, og hvilke der taler for ham.

11: afgør ved at beregne scoren om man vil tænke at en mail er ham eller spam i følgende 2 situationer

a) Oprindelse : andet, sendt fra hotmail-konto og mail omhandler ikke dating eller gevinst.

b) Mail fra DK, firma-mail med indhold relateret til dating