

1 of 1

118.95%

Thumbnails

1

2 of 10

155.37%

Question 1

a) Some emails received by users are spams, attempting to gain illicit access to employee's accounts. You are building a classifier to detect such illicit emails. You start by using the feature of whether or not an email asks for the recipients password, as this an important datum indicating if the email is in fact a hacking attempt. In particular 95% of hacking emails contain requests for a user's password, and 95% of legitimate emails do not contain requests for a user's password. 1% of emails overall are hacking attempts.

- If your classifier scans an email with a password request, what is the probability that the email is in fact a hacking attempt?
- Comment on this value that you calculate.
- What is the probability that your classifier makes an error?

$p(H)=0.01$. $P(P|H)=0.95$. $P(!P|H)=0.95$.

$\Rightarrow p(!H) = 1-p(H) = 0.99$. $p(P|!H) = 1-p(!P|!H) = 1-0.95 = 0.05$

$p(H|P) = \frac{p(P|H)p(H)}{P(P)} = \frac{p(P|H)p(H)}{p(P|H)p(H)+p(P|!H)p(!H)}$
 $= \frac{0.95*0.01}{0.95*0.01 + 0.05*0.99} = \frac{0.0095}{0.0095 + 0.0495} = \frac{0.0095}{0.059} = 0.161$.

Four marks for correct answer, three for working.

The correct detection probability is low due to the tiny prior. But it's better than relying on the prior alone, which would be 0.01.