Exercise 5

Notations

S = message is actually spam
L = message is actually legitimate
$\hat{S}$ = message is classified as spam
$\hat{L}$ = message is classified as legitimate

Part (a)

From the given data, we conclude the following:

- $\Pr\left(\hat{S} \mid S\right) = 0.85$
- $\Pr\left(\hat{S} \mid L\right) = 0.05$
- $\Pr\left(S\right) = 0.60$
- $\Pr\left(L\right) = 1 - \Pr\left(S\right) = 1 - 0.60 = 0.40$
- $\Pr\left(\hat{S} \cap S\right) = \Pr\left(\hat{S} \mid S\right) \Pr\left(S\right) = 0.85 \times 0.60 = 0.51$
- $\Pr\left(\hat{S} \cap L\right) = \Pr\left(\hat{S} \mid L\right) \Pr\left(L\right) = 0.05 \times 0.40 = 0.02$
- $\Pr\left(\hat{S}\right) = \Pr\left(\hat{S} \cap S\right) + \Pr\left(\hat{S} \cap L\right) = 0.51 + 0.02 = 0.53$
- $\Pr\left(\hat{L} \cap S\right) = \Pr\left(S\right) - \Pr\left(\hat{S} \cap S\right) = 0.60 - 0.51 = 0.09$
- $\Pr\left(\hat{L} \cap L\right) = \Pr\left(L\right) - \Pr\left(\hat{S} \cap L\right) = 0.40 - 0.02 = 0.38$

	S	L	
$\hat{S}$	0.51	0.02	0.53
$\hat{L}$	0.09	0.38	0.47
	0.60	0.40	1

From the above table, we can conclude the following[1]:

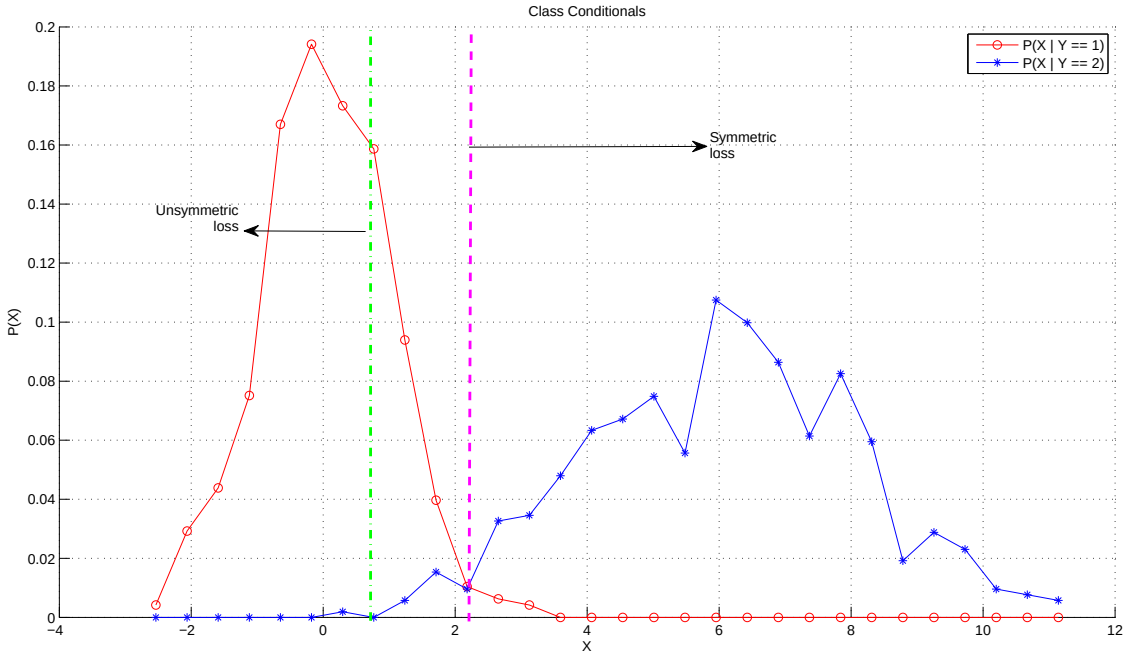
- False positive $= \Pr\left(\hat{S} \mid L\right) = \frac{\Pr\left(\hat{S} \cap L\right)}{\Pr(L)} = \frac{0.02}{0.40} = 0.05$
- False negative $= \Pr\left(\hat{L} \mid S\right) = \frac{\Pr\left(\hat{L} \cap S\right)}{\Pr(S)} = \frac{0.09}{0.60} = 0.15$
- Average error $= \Pr\left(\hat{S} \cap L\right) + \Pr\left(\hat{L} \cap S\right) = 0.02 + 0.09 = 0.11$

Part (b)

- A classication algorithm with zero false positive rate would be a classifier which simply classifies everything as legitimate.
- A classication algorithm with zero false negative rate would be a classifier which simply classifies everything as spam.

[1] Definitions from "Counting performance measures (2)" page of the slides

Part (c)


Class Conditionals
P(X | Y == 1)
P(X | Y == 2)
Symmetric loss
Unsymmetric loss
P(X)
X
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
−4
−2
0
2
4
6
8
10
12