Q2 (a)

word	P(word | spam) / P(word | ¬spam)	P(¬word | spam) / P(¬word | ¬spam)
w1	0.8/0.2 = 4	0.2/0.8 = 1/4
w2	0.5/0.5 = 1	0.5/0.5 = 1
w3	0.1/0.4 = 1/4	0.9/0.6 = 3/2

(b) An email containing w1 but not w3 is maximally likely to be spam (LR 6). An email containing w3 but not w1 is maximally likely to be non-spam (LR 1/16). The presence or absence of w2 is immaterial.
If spam is 10 times more likely to occur than non-spam, then the first email is still predicted as spam (10*6 > 1) and the second is still predicted as non-spam (10*1/16 < 1).

(c) The worst feature to split on is w2: w2 present: (5 spam, 5 non-spam); w2 absent: (5 spam, 5 non-spam). Both subsets have entropy 1, so clearly zero information gain.
w1 present: (8 spam, 2 non-spam); w1 absent: (2 spam, 8 non-spam). Both subsets have the same entropy, say E.
w3 present: (1 spam, 4 non-spam), again entropy E; w3 absent: (9 spam, 6 non-spam), with an entropy higher than E.
So the best feature is presence/absence of w1.

(d) In general we could use this information to construct a new feature testing presence/absence of both w1 and w2. However, the numbers indicate that w1 and w2 are independent given the class (spam: 8*5/10 = 4; non-spam: 2*5/10 = 1), so this new feature would not change classification accuracy.