

Homework 11
Jake Schinto

a. Predicted
 True False

Actual			
True	90	10	
False	60	40	

b. It corresponds to expected profit because as stated in the passage before the curves, "For that purpose we use a profit curve, which incorporates assumptions about costs and benefits and displays expected value." Therefore the profit curves show the expected values for profit and can be used to help guide obtaining maximum overall profit.

c. You would use additive smoothing or the Laplace Correction to correct the $p(e_i)$. So, $1/1000 \Rightarrow (1+1)/(1000+1) = (2/1001) = 0.001998$ instead of the original $.001$

d. 1000 spam and 9000 not spam
 $\text{lift}(x) = p(x|c)/p(x) = p(\text{"free"}|\text{spam})/p(\text{"free"})$
 $p(\text{"free"}|\text{spam}) = 50/1000 = .05$
 $p(\text{"free"}) = 100/10000 = .01$
 $\text{lift}(\text{"free"}) = .05/.01 = 5$

e. $\text{IDF}(X) = 1 + \log(200 / 20) = 1 + \log(10) = 1 + 1 = 2$