

Ex 2:

1.

$$P(\text{million} | \text{spam}) = \frac{156}{95791}$$

$$P(\text{dollars} | \text{spam}) = \frac{29}{95791}$$

$$P(\text{adclick} | \text{spam}) = \frac{51}{95791}$$

$$P(\text{conferences} | \text{spam}) = \frac{0}{95791} \approx 0.000001$$

$$P(\text{million} | \text{ham}) = \frac{98}{95791}$$

$$P(\text{dollars} | \text{ham}) = \frac{119}{95791}$$

$$P(\text{adclick} | \text{ham}) = \frac{0}{95791} \approx 0.000001$$

$$P(\text{conferences} | \text{ham}) = \frac{12}{95791}$$

$$2. \quad P(\text{Word} \neq \text{million}) = 1 - P(\text{Word} = \text{million}) = 1 - (0.5 * P(\text{million} | \text{spam}) + 0.5 * P(\text{million} | \text{ham})) = 1 - (0.5 * \frac{156}{95791} + 0.5 * \frac{98}{95791}) = 1 - \frac{127}{95791} = \frac{95664}{95791} \approx 0.9987$$

$$3. \quad P(\text{spam} | \text{million}) = \frac{P(\text{million} | \text{spam})P(\text{spam})}{P(\text{million})} = \frac{\frac{156}{95791} * \frac{1}{2}}{\frac{1}{2} * \frac{95664}{95791}} = \frac{78}{127} \approx 0.6142$$

$$4. \quad P(\text{spam} | \text{million, dollars, adclick, conferences}) = \frac{P(\text{million, dollars, adclick, conferences} | \text{spam})P(\text{spam})}{P(\text{million, dollars, adclick, conferences})} = \frac{P(\text{million} | \text{spam})P(\text{dollars} | \text{spam})P(\text{adclick} | \text{spam})P(\text{conferences} | \text{spam})P(\text{spam})}{P(\text{million})P(\text{dollars})P(\text{adclick})P(\text{conferences})} = \frac{P(\text{million} | \text{spam})P(\text{dollars} | \text{spam})P(\text{adclick} | \text{spam})P(\text{conferences} | \text{spam})P(\text{spam})}{(P(\text{million} | \text{spam}) + P(\text{million} | \text{ham}) + P(\text{dollars} | \text{spam}) + P(\text{dollars} | \text{ham}) + P(\text{adclick} | \text{spam}) + P(\text{adclick} | \text{ham}) + P(\text{conferences} | \text{spam}) + P(\text{conferences} | \text{ham}))} = \frac{\frac{156}{95791} * \frac{29}{95791} * \frac{51}{95791} * 0.000001 * \frac{1}{2}}{\frac{1}{2} (\frac{156}{95791} + \frac{29}{95791} + \frac{51}{95791} + 0.000001 + \frac{98}{95791} + \frac{119}{95791} + 0.000001 + \frac{12}{95791})} \approx 5.4 * 10^{-14}$$

It does not appear to be spam, but it's also clearly an odd case with both the word conferences and adclick.

$$\text{spamcity} = \frac{P(\text{spam} | \text{million, dollars, adclick, conferences})}{P(\text{ham} | \text{million, dollars, adclick, conferences})} = \frac{P(\text{spam}) P(\text{million} | \text{spam}) P(\text{dollars} | \text{spam}) P(\text{adclick} | \text{spam}) P(\text{conferences} | \text{spam})}{P(\text{ham}) P(\text{million} | \text{ham}) P(\text{dollars} | \text{ham}) P(\text{adclick} | \text{ham}) P(\text{conferences} | \text{ham})} = \frac{\frac{1}{2} \frac{156}{95791} \frac{29}{95791} \frac{51}{95791} 0.000001}{\frac{1}{2} \frac{98}{95791} \frac{119}{95791} 0.000001 \frac{12}{95791}} \approx 1.6$$