Q4

Bayes theorem: $\mathrm{P(A|B)} = \frac{P(B|A)P(A)}{P(B)}$

From the question:

P(Red) = 0.6

P(Blue) = 0.4

P(Apple|Red) = 0.25

P(Orange|Red) = 0.75

P(Apple|Blue) = 0.75

P(Orange|Blue) = 0.25

P(Blue|Orange) $= \frac{P(Orange|Blue)P(Blue)}{P(Orange)}$

$= \frac{0.25*0.4}{0.75*0.6+0.25*0.4}$

$= \frac{2}{11}$

Q5

Batch gradient descent

The objective function of batch gradient descent for linear regression with L2 regularization is:

$$\frac{1}{2}\sum_{n=1}^{N}(t_n - w \bullet \emptyset(x_n)^2) + \frac{\lambda}{2}\sum_{j=0}^{M-1} w_j^2$$

The gradient for the above function is

$$\sum_{n=1}^{N}\big(t_n - w \bullet \emptyset(x_n)\big)\big(-\emptyset(x_n)\big) + w\lambda$$

Therefore, we need to update each of the w, by multiplying the gradient by – η and add this term to the original w. For example, for updating w_1:

$$w_{1,n+1} = w_{1,n} + \eta\sum_{n=1}^{N}\big(t_n - w \bullet \emptyset(x_n)\big)\big(\emptyset(x_n)\big) - \eta\lambda w_1$$

The only difference between batch gradient descent and stochastic gradient descent is that, the weight is updated after visiting each data point rather than the whole training set. Therefore the objective function is:

$$\frac{1}{2}(t_n - w \bullet \emptyset(x_n)^2) + \frac{\lambda}{2}\sum_{j=0}^{M-1} w_j^2$$

And the gradient is:

$$\big(t_n - w \bullet \emptyset(x_n)\big)\big(-\emptyset(x_n)\big) + w\lambda$$

The update for w_1 is:

$$w_{1,n+1} = w_{1,n} + \eta\big(t_n - w \bullet \emptyset(x_n)\big)\big(\emptyset(x_n)\big) - \eta\lambda w_1$$