Q 6

(a)

e ^ (b0 + b1*x1 + b2*x2) / (1 + e ^ (b0 + b1*x1 + b2*x2)) =
e ^ (-6 + 0.05 *40 + 1*3.5) / (1 + e ^ (-6 + 0.05 *40 + 1*3.5)) = 0.3775 **(37.75%)**

(b)

we have to solve the equation: (x1 = hours)
e ^ (-6 + 0.05 *x1 + 1*3.5) / (1 + e ^ (-6 + 0.05 *x1 + 1*3.5)) = 0.5
let's assume y = e ^ (-6 + 0.05 *x1 + 1*3.5)
y / (1+y)=0.5 => y = 1
(y =)e ^ (-6 + 0.05 *x1 + 1*3.5) = 1
-6 + 0.05 *x1 + 1*3.5 = 0
x1 = **50h**

Q 7

- Plug into Bayes:

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)}$$

- Two tricks

solution in attached image q7.png
final answer: 74.13%

Q8

For KNN with K=1, the training error rate is 0% . So, KNN has a test error rate of 18% * 2 = 36%
We shouldp prefer logistic regression, because it has lower error rate (30%)

Q 12

(a)

$$\log\left(\frac{p_1}{1-p_1}\right) = \beta_0 + \beta_1 x.$$

(b) (alpha_orange0 - alpha_apple0) + (alpha_orange1 - alpha_apple1)*x

$$\log\left(\frac{\Pr(Y=k|X=x)}{\Pr(Y=k'|X=x)}\right) = (\beta_{k0} - \beta_{k'0}) + (\beta_{k1} - \beta_{k'1})x_1 +$$

(c)

(d)
a_orange0 =1.2, a_orange1 = -2, a_apple0 =3, a_apple1=0.6

I belive there is an error in the book! a_apple referred as an a_orange

according to (b): beta0 = 1.2 – 3 = -1.8, beta1 = -2 – 0.6 = -2.6

(e)
100% of time. The same model.

Q 5 (chapter 5)

(a)
The training set is split into *k* smaller sets.
The following procedure is followed for each of the *k* “folds”:

- A model is trained using k-1 of the folds as training data;
- the resulting model is validated on the remaining part of the data (i.e., it is used as a test set to compute a performance measure such as accuracy).

(b)

When using a validation set, we can only train on a small portion of the data as we must reserve the rest for validation. As a result it can overestimate the test error rate (assuming we then train using the complete data for future prediction). It is also sensitive to which observations are including in train vs. test. It is, however, low cost in terms of processing time (as we only have to fit one model).

When using LOOCV, we can train on n-1 observations, however, the trained models we generate each differ only by the inclusion (and exclusion) of a single observation. As a result, LOOCV can have high variance (the models fit will be similar, and might be quite different to what we would obtain with a different data set). LOOCV is also costly in terms of processing time.