1.

   a.

      i. Given there are 5 different reviewer and 10 possible scores each reviwer could give (1-10), this is a case of permuation with repetition (as each reviewer can choose the same score if they want, and the order of scores matters). The formula for this is n^r where n is the number of objects and r is the number of times a selection is made.
      This means there are 10^5 combinations of ratings, or $100,000$

      ii. Here there are 8 people being selected from 25, and the order does not matter (as selecting X and Y to attend is the same as selecting Y and X). However, the same person cannot be selected more than once. The formula for combination without repetition is $\binom{n}{r} = \frac{n!}{(n-r)!r!}$.
      So there are $\frac{25!}{(25-8)!8!} = 1,081,575$ different groups of 8 people possible at movie night.

      iii. Putting 500 pieces of popcorn into 3 different bowls is combination with repetition, and so the formula is $\binom{r+(n-1)}{n-1}$.
      However, there are not 500 objects here, as you can’t have a bowl without any popcorn! First, place a piece of popcorn in each bowl, leaving 497. There are then $\binom{497+(3-1)}{3-1} = \binom{499}{2} = \frac{499!}{(499-2)!2!} = 124,251$ different ways to divide the popcorn between bowls. Just make sure to keep the fullest one for yourself!

      iv. This is an example of conditional probability – what’s the probability of watching a horror movie given the movie selected is positive. The formula $P(B)$ for this is $P(A|B) = \frac{P(A \cap B)}{P(B)}$. So if we let watching a horror $= 0.417$ movie = A and watching a movie with positive review = B, we then need to understand $P(B)$ and $P(A \cap B)$. $P(B)$ is just the number of positive movies (11) divided by the number of movies (24). So $P(B) = \frac{11}{24}$. Similarly, $P(A)$ is the number of horror movies (10) divided by the total movies, or $P(B) = \frac{10}{24}$. Now, $P(A \cap B)$ is just the probability of both $P(A)$ and $P(B)$ both occurring. This means $P(A \cap B) = \frac{11}{24} \times \frac{10}{24} = \frac{55}{288} = 0.191$. We can now use the formula, and see that $P(A|B) = \frac{0.191}{0.458} = 0.417$. This is the probability that you’re watching a horror moive if you choose a movie to watch at random and it received positive reviews.

   b.

i. Calculating the probability of each word is simply the number of times the word occurs divided by the total number of reviews (1000 positive +1000 negative = 2000). As there are equal number of positive and negative reviews, the probability of a random word coming from a positive review or from a negative review is $\frac{1000}{2000}$ or $\frac{1}{2}$. Taking the word 'recommend' as an example, with 81 occurrences in positive words, the conditional probability of the word occurring in a positive word is given by $\frac{81}{2000} \div \frac{1}{2}$ - using $P(A|B)\frac{P(A \cap B)}{P(B)}$. This can actually be simplifed to $P(w|s) = \frac{n}{1000}$ where P(w) is the probabiltiy of the word, P(S) is the probability of the sentiment (positive or negative) and n is the number of times that word occurs within a review of sentiment S.

Word	P(w|Pos)	P(w|Neg)	P(w)
Recommend	0.081	0.057	0.069
Hilarous	0.062	0.019	0.041
Obvious	0.034	0.062	0.048
Problems	0.031	0.03	0.031
Awkward	0.008	0.006	0.007
Boredom	0	0.012	0.006

ii. The probability a review is negative given that it contains the word 'awkward' is given by $\frac{P(neg \cap w)}{p(w)}$ were p(w) is the probability of the word (in this case awkward) and P(neg) is the probability of a review being negative). This means the probability that a review is negative given that it contains 'awkward' is $\frac{0.003}{0.007} = 0.429$

iii. Bayes theorum is $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$. Since we're trying to determine the probability of a review being negative P(Neg) given it includes 'hilarious', 'obvious' and 'problems', we can define this as $P(neg|h, o, b)$ where P(h) is the probability of 'hilarious' occuring, P(o) is the probability of 'obvious' occuring and P(b) is the probability of 'problems' occuring. Using $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$, we can now find $P(neg|h, o, b)$ by using $\frac{P(h|neg)P(o|neg)P(b|neg)P(neg)}{P(h, o, b)}$.
Using $P(A|B)\frac{P(A \cap B)}{P(B)}$, we can simplify this to $\frac{P(h|neg)P(o|neg)P(b|neg)P(neg)}{P(h, o, b)} = \frac{0.019 \cdot 0.062 \cdot 0.03 \cdot 0.5}{0.041 \cdot 0.48 \cdot 0.031} = \frac{0.00001767}{0.00020832} = 0.085$. This means the probability of a review containing the words 'hilarious','obvious' and 'problems' being negative is 0.085. We

then use the same approach to determine the probability of the same review being positive. $P(pos|h,o,b) = \frac{P(h|posP(o|pos)P(b|pos)P(pos)}{P(h,o,b)} = \frac{0.062 \cdot 0.034 \cdot 0.031 \cdot 0.5}{0.041 \cdot 0.048 \cdot 0.031}$ $=0.536$. This means the probability of the review being postive is higher, so the review is more likely to be positive if it includes the words 'hilarious','obvious' and 'problems'.

iv. Using the same approach as in the previous question, the probability that the review containing the words 'boredom' and 'recommend' being positive is $\frac{0 \cdot 0.081 \cdot 0.5}{0.006 \cdot 0.069} = 0$. The probability of the review being negative is $\frac{0.012 \cdot 0.057 \cdot 0.5}{0.006 \cdot 0.069} = \frac{0.000342}{0.000414} = 0.826$. This is obviously greater than 0, so the review is more likely to be negative.

v. One problem is that if a word doesn't appear in the 'training' or 'sample' data, in a certain sentiment class, the formula will never predict it occurring in that kind of review. Seen here, as there are no cases of 'boredom' appearing in a positive review in the training data, it will always give the probability of any new review which includes that word being postive at 0. Another issue is that it assumes the conditional independence of the events – i.e that the probability of a review containing 'boredom' is independent from the probability of the same review containing 'recommend'. In practice, many words are more likely to occur together, and this may affect the usefulness of the Naïve Bayes results.

vi. The new probabilities of 'boredom' and 'recommend' after Laplace smoothing are:

Word	P(w|Pos)	P(w|Neg)	P(w)
Recommend	0.082	0.058	0.070
Boredom	0.001	0.013	0.007

This means the probability of a review containing this words being positive is now $\frac{0.001 \cdot 0.082 \cdot 0.5}{0.007 \cdot 0.069} = \frac{0.000041}{0.000483} = 0.009$. The probability of it being negative is $\frac{0.013 \cdot 0.058 \cdot 0.5}{0.007 \cdot 0.070} = \frac{0.000377}{0.00049} = 0.769$. It is still more likely to be negative, but now the Naïve Bayes approach presents some possibility it is positive.

2.

a.

i. As $X = \begin{bmatrix} 1\,0\,0 \\ 1\,3\,9 \\ 1\,6\,36 \end{bmatrix}$, then $X^T = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 3 & 6 \\ 0 & 9 & 36 \end{bmatrix}$. So

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 3 & 6 \\ 0 & 9 & 36 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & 9 \\ 1 & 6 & 36 \end{bmatrix} = \begin{bmatrix} 3 & 9 & 45 \\ 9 & 45 & 243 \\ 45 & 243 & 1377 \end{bmatrix}$$

ii. Given the value of $(X^T X)$ found in the previous questions, then

$$(X^T X)^{-1} = \begin{bmatrix} 3 & 9 & 45 \\ 9 & 45 & 243 \\ 45 & 243 & 1377 \end{bmatrix}^{-1} = \begin{bmatrix} 3 & 9 & 45 \\ 9 & 45 & 243 \\ 45 & 243 & 1377 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

. Using row operations:

$R_2 = R_2 - 3R_1$:

$$\begin{bmatrix} 3 & 9 & 45 \\ 0 & 18 & 108 \\ 45 & 243 & 1377 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$R_1 = R_1 \div 3$:

$$\begin{bmatrix} 1 & 3 & 15 \\ 0 & 18 & 108 \\ 45 & 243 & 1377 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$R_3 = R_3 - 45R_1$:

$$\begin{bmatrix} 1 & 3 & 15 \\ 0 & 18 & 108 \\ 0 & 108 & 702 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ -3 & 1 & 0 \\ -15 & 0 & 1 \end{bmatrix},$$

$R_2 = \dfrac{R_2}{18}$:

$$\begin{bmatrix} 1 & 3 & 15 \\ 0 & 1 & 6 \\ 0 & 108 & 702 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ -\frac{3}{18} & \frac{1}{18} & 0 \\ -15 & 0 & 1 \end{bmatrix},$$

$R_1 = R_1 - 3R_2$:

$$\begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 6 \\ 0 & 108 & 702 \end{bmatrix} \begin{bmatrix} \frac{5}{6} & -\frac{1}{6} & 0 \\ -\frac{3}{18} & \frac{1}{18} & 0 \\ -15 & 0 & 1 \end{bmatrix},$$

$R_3 = R_3 - 108R_2$:

$$\begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 6 \\ 0 & 0 & 54 \end{bmatrix} \begin{bmatrix} \frac{5}{6} & -\frac{1}{6} & 0 \\ -\frac{3}{18} & \frac{1}{18} & 0 \\ 3 & -6 & 1 \end{bmatrix},$$

$R_3 = \frac{R_3}{54}$:

$$\begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{5}{6} & -\frac{1}{6} & 0 \\ -\frac{3}{18} & \frac{1}{18} & 0 \\ \frac{3}{54} & -\frac{6}{54} & \frac{1}{54} \end{bmatrix}$$,

$R_2 = R_2 - 6R_3$:

$$\begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{5}{6} & -\frac{1}{6} & 0 \\ -\frac{9}{18} & \frac{13}{18} & -\frac{1}{9} \\ \frac{3}{54} & -\frac{6}{54} & \frac{1}{54} \end{bmatrix}$$,

$R_1 = R_1 + 3R_3$:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & \frac{1}{18} \\ -\frac{9}{18} & \frac{13}{18} & -\frac{1}{9} \\ \frac{3}{54} & -\frac{6}{54} & \frac{1}{54} \end{bmatrix}$$ simplified, means that

$$\left(X^T X\right)^{-1} = \begin{bmatrix} 1 & -\frac{1}{2} & \frac{1}{18} \\ -\frac{1}{2} & \frac{13}{18} & -\frac{1}{9} \\ \frac{1}{18} & -\frac{1}{9} & \frac{1}{54} \end{bmatrix}$$

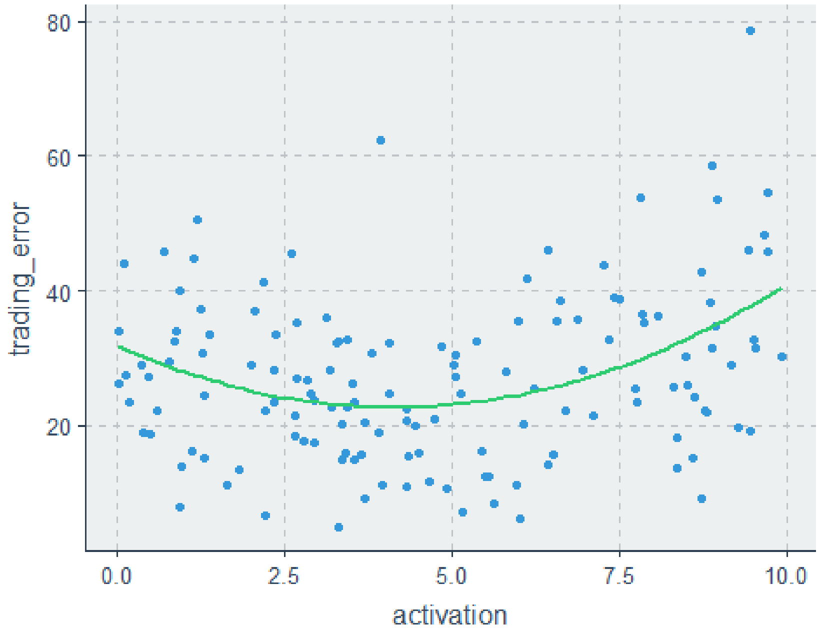
iii. The formula for calculating the regression coefficient is $\hat{\beta} = \left(X^T X\right)^{-1} X^T y$. As $\left(X^T X\right)^{-1} = \begin{bmatrix} 1 & -\frac{1}{2} & \frac{1}{18} \\ -\frac{1}{2} & \frac{13}{18} & -\frac{1}{9} \\ \frac{1}{18} & -\frac{1}{9} & \frac{1}{54} \end{bmatrix}$, it has order 3x3. We also know X^T has order 3x3, so $\left(X^T X\right)^{-1} X^T$ will have order 3x3. From the given points, we know y is a 3x1 matrix, and so $\hat{\beta}$ will have order 3x1 and = $\begin{pmatrix} 34 \\ -9 \\ \frac{10}{9} \end{pmatrix}$.

iv. Using the equation for a quadratic ($y=ax^2 + bx + c$), the quadratic for this is $y = \frac{10}{9}x^2 - 9x + 34$. You can see here there is no error term – this is because the coefficient was produced by minimising the mean squre error to produce a mean of zero. On average, there is no error in the y produced, meaning no error can be added as standard.

b.

i. As per appendix A, the line of best fit is $y = 0.533x^2 - 4.42x + 31.912$.

ii. Scatter plot of data with line of best fit :



iii. Using the line of best fit, if y = 25, then $0.533x^2 -4.42x + 31.912 = 25$ this means that $0.533x^2 -4.42x + 6.912= 0$. Using the solutions to a quadratic equation $x_{1,\,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, there will be two possible values of x:

$$x_{1,\,2} = \frac{4.42 \pm \sqrt{-4.42^2 - 4 * 0.533 * 6.912}}{2 * 0.533} =$$

$$x_{1,\,2} = \frac{4.42 \pm \sqrt{19.5364 - 14.736384}}{1.066} =$$

$$x_{1,\,2} = \frac{4.42 \pm \sqrt{4.8}}{1.066}$$

. As there are two possible versions of X at this point, there is no inverse for this line.

Appendix A

Kristen Osborne

```
#create a new variable for trading_error
raw_data$activation_sq <- raw_data$activation^2

#fit quadratic regression model
quadratic_model <- lm(trading_error ~ activation + activation_sq, da
ta=raw_data)

#view model summary
summary(quadratic_model)
```

```
##
## Call:
## lm(formula = trading_error ~ activation + activation_sq, data = raw_
data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.600  -7.671  -0.999   8.555  40.876
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    31.9124     2.8382  11.244  < 2e-16 ***
## activation     -4.4299     1.3370  -3.313  0.00118 **
## activation_sq   0.5334     0.1298   4.110 6.75e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.63 on 138 degrees of freedom
## Multiple R-squared:  0.1471, Adjusted R-squared:  0.1347
## F-statistic:  11.9 on 2 and 138 DF,  p-value: 1.708e-05
```

```
ggplot(raw_data, aes(activation,trading_error,activation_sq)) + geom
_point()+stat_smooth(se=F, method='lm', formula=y ~ x + I(x^2))
```