Lawrence Ouyang
504128219
Math 191
Homework 2

1a.

$$f_X(x) = \begin{cases} \dfrac{1}{2}, & -1 \le x \le 1 \\ \\ 0, & x < -1, x > 1 \end{cases}$$

$$f_Y(y) = \begin{cases} 1, & 0 \le y \le 1 \\ \\ 0, & y < 0, y > 1 \end{cases}$$

$$f_{Y|X}(y|x) = \begin{cases} 1, & -1 \le x < 0 \\ 1, & 0 < x \le 1 \end{cases}$$

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x) = f_X(x)f_Y(y)$$

Thus X and Y are independent and their correlation is 0.

1b. Since α is distributed uniformly $\alpha \sim \text{Unif}([0, 2\pi])$, then X and Y are uniformly distributed on $[-1, 1]$ and $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. Thus $cov(X,Y) = cor(X,Y) = 0$ which means X and Y are uncorrelated. However, by trigonometric identities, we notice that for all $x \in X, y \in Y = \pm\sqrt{1-x^2}$, thus Y is dependent of X.

2a. The closed form for $\hat{\beta}_0$ is $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$. We denote $\bar{y} = \frac{1}{n}\sum_{i=1}^n \beta_0 + \beta_1 x_i + e_i$ and $\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$:

$$\begin{aligned} \mathbb{E}[\hat{\beta}_0] &= \mathbb{E}[\bar{y}] - \bar{x}\mathbb{E}[\hat{\beta}_1] \\ &= \mathbb{E}[\frac{1}{n}\sum_{i=1}^n \beta_0 + \beta_1 x_i + e_i] - \bar{x}\hat{\beta}_1 \\ &= \beta_0 + \hat{\beta}_1\bar{x} - \bar{x}\hat{\beta}_1 \\ &= \beta_0 \end{aligned}$$

2b. Consider the OLS Estimator $\hat{\beta}$ to have the form $(X^TX)^{-1}X^Ty$, where $y = X\beta + \mu$ and μ is the residual.

$$\begin{aligned} \mathbb{E}[\hat{\beta}] &= \mathbb{E}[(X^TX)^{-1}X^Ty] \\ &= \mathbb{E}[(X^TX)^{-1}X^T(X\beta + \mu)] \\ &= \mathbb{E}[(X^TX)^{-1}X^TX\beta + (X^TX)^{-1}X^T\mu] \\ &= \mathbb{E}[\beta + (X^TX)^{-1}X^T\mu] \\ &= \beta + \mathbb{E}[(X^TX)^{-1}X^T\mu], \mathbb{E}[\mu] = 0 \\ &= \beta \end{aligned}$$

3. Use the same notations from 2a.

$$\begin{aligned}
var(\hat{\beta}_0) &= var(\bar{y} - \hat{\beta}_1\bar{x}) \\
&= var(\bar{y}) - \bar{x}^2 var(\hat{\beta}_1) - 2cov(\bar{y}, \bar{x}\hat{\beta}_1) \\
&= \frac{var(y_i)}{n} - \bar{x}^2 \frac{var(y_i)}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} - 2\bar{x}^2 cov(\bar{y}, \hat{\beta}_1), var(y_i) = \sigma^2 \\
&= \frac{\sigma^2}{n} - \frac{\bar{x}^2\sigma^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} - 2\bar{x}^2 cov(\bar{y}, \hat{\beta}_1) \\
cov(\bar{y}, \hat{\beta}_1) &= \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} cov(\bar{y}, y_i) \\
&= 0 \\
var(\hat{\beta}_0) &= \frac{\sigma^2}{n} - \frac{\bar{x}^2\sigma^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \\
&= \sigma^2 \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2 + n\bar{x}^2}{n\sum_{i=1}^n x_i^2 - n\bar{x}^2} \\
&= \sigma^2 \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - n\bar{x}^2}
\end{aligned}$$

4.

$$\phi_X(t) = \begin{cases} \mathbb{E}[e^{itX}], & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

$$\begin{aligned}
\phi_X(t) &= \mathbb{E}[e^{itX}] \\
&= \int_0^\infty \lambda e^{-\lambda x} e^{itx} dx \\
&= \lambda \int_0^\infty e^{x(it-\lambda)} dx \\
&= \frac{\lambda}{it - \lambda} e^{x(it-\lambda)} \Big|_0^\infty \\
&= \frac{\lambda}{it - \lambda}
\end{aligned}$$

5.

$$\begin{aligned}
H(X,Y) &= -\sum_x \sum_y p(x,y) log p(x,y) \\
&= -\sum_x \sum_y p(x,y) log(p(x)p(y|x)) \\
&= -\sum_x \sum_y p(x,y) log p(x) - \sum_x \sum_y p(x,y) log p(y|x) \\
&= -\sum_x p(x) log p(x) - \sum_x \sum_y p(x,y) log p(y|x) \\
&= H(X) + H(Y|X)
\end{aligned}$$

6a. Marginal distributions:

$$p_x(X) = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$$
$$p_y(Y) = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$$

6b. Entropy:

$$H(X) = -\frac{1}{2}log\frac{1}{2} - \frac{1}{4}log\frac{1}{4} - 2\frac{1}{8}log\frac{1}{8} = \frac{7}{4}$$
$$H(Y) = -log\frac{1}{4} = 2$$

6c. Joint/Conditional Entropy:

$$\begin{aligned}
H(X|Y) &= \sum_{i=1}^{4} p(Y=i)H(X|Y=i) \\
&= \frac{1}{4}(H(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}) + H(\frac{1}{4}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8})) + H(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}) + H(1,0,0,0)) \\
&= \frac{1}{4}(\frac{7}{4} + \frac{7}{4} + 2) = \frac{11}{8}
\end{aligned}$$

$$\begin{aligned}
H(Y|X) &= \sum_{i=1}^{4} p(X=i)H(Y|X=i) \\
&= \frac{1}{2}H(\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{2}) + \frac{1}{4}H(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0) + \frac{1}{4}H(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0) \\
&= \frac{1}{2}(\frac{7}{4}) + \frac{1}{2}(\frac{3}{2}) = \frac{13}{8}
\end{aligned}$$

$$\begin{aligned}
H(X,Y) &= H(X) + H(Y|X) \\
&= \frac{7}{4} + \frac{13}{8} = \frac{27}{8}
\end{aligned}$$

7a. Consider mutual information to be defined as:

$$\begin{aligned}
I(X;Y) &= \sum_x \sum_y p(x,y)log\frac{p(x,y)}{p(x)p(y)} \\
&= \sum_x \sum_y p(x,y)log\frac{p(x|y)}{p(x)} \\
&= \sum_x \sum_y p(x,y)logp(x|y) - \sum_x \sum_y p(x,y)logp(x) \\
&= -H(X|Y) + H(X) \\
&= H(X) - H(X|Y)
\end{aligned}$$

7b.

$$
\begin{aligned}
I(X;Y) &= \sum_x \sum_y p(x,y) log \frac{p(x,y)}{p(x)p(y)} \\
&= \sum_x \sum_y p(x,y) log p(x,y) - \sum_x \sum_y p(x,y) log p(x) - \sum_x \sum_y p(x,y) log p(y) \\
&= \sum_x \sum_y p(x,y) log p(x,y) - \sum_x p(x) log p(x) - \sum_y p(y) log p(y) \\
&= -H(X,Y) + H(X) + H(Y) \\
&= H(X) + H(Y) - H(X,Y)
\end{aligned}
$$

7c.

$$
\begin{aligned}
I(X;Y) &= \sum_x \sum_y p(x,y) log \frac{p(x,y)}{p(x)p(y)} \\
&= \sum_x \sum_y p(x,y) log \frac{p(y|x)}{p(y)} \\
&= \sum_x \sum_y p(x,y) log p(y|x) - \sum_x \sum_y p(x,y) log p(y) \\
&= -H(Y|X) + H(Y) \\
&= H(Y) - H(Y|X) \\
&= I(Y;X)
\end{aligned}
$$

7d.

$$
\begin{aligned}
I(X;X) &= \sum_x \sum_x p(x,x) log \frac{p(x,x)}{p(x)p(x)} \\
&= \sum_x p(x) log \frac{p(x)}{p(x)} - \sum_x p(x) log p(x) \\
&= -\sum_x p(x) log p(x) \\
&= H(X)
\end{aligned}
$$

8ai. Yes, as the predictor(horsepower) increases, the response(mpg) decreases. The coefficients given for this simple linear regression model is 39.935861 for the intercept and -0.157845 for the slope.

8aii. The relationship is somewhat strong with an adjusted R-squared score of 0.6049. The near zero p-value does indicate a strong rejection of the null.

8aiii. The relationship is negative: as the predictor increases, the response decreases.

8aiv. The predicted mpg associated with 98 horsepower is 24.46708, with a confidence interval [23.97308, 24.96108] and prediction interval [14.8094, 34.12476].

8b, 8c. See attached plots.

9. Programming Assignment
(a). Using the order {SPY, XVIX, TNX, OIL, GOLD, N225, FTSE, GLD}, our β values are:

$$\beta_0 = 0.0004742$$
$$\beta_1 = -0.1576429$$
$$\beta_2 = -0.0067277$$
$$\beta_3 = 0.0129053$$
$$\beta_4 = 0.0030628$$

$$\beta_5 = 0.0003391$$
$$\beta_6 = 0.0106196$$
$$\beta_7 = 0.0930072$$
$$\beta_8 = 0.0409372$$

$$\begin{aligned}\hat{r}_{t+1,SPY} &= 0.0004742 - 0.1576429r_{t,SPY} - 0.0067277r_{t,XVIX} + 0.0129053r_{t,TNX} + 0.0030628r_{t,OIL} \\ &+ 0.0003391r_{t,GOLD} + 0.0106196r_{t,N225} + 0.0930072_{t,FTSE} + 0.0409372_{t,GLD}\end{aligned}$$

(b). Here is the code for part b:

```
doWork_In_Sample = 1
if(doWork_In_Sample == 1){

   X_train = RET[1:nrDays - 1,];
   X_test = RET;
   STATS = NULL;
   CM = NULL;
   for (i in 1:nrTickers) {
      print('###############################################');
      print( paste('Stock = ',tickers[i] ) );
      y_train = RET[ 2 : nrDays , tickers[i], drop=FALSE]; # "drop=FALSE" ensurs it remains
         a 2-D array..

      y_hat = compute_linear_regression(X_train, y_train, X_test);
      # compute the various performance statistics:
```

```
    daily_pnl = matrix(, nrow = nrDays - 1, ncol = 1);
    for (j in 1:(nrDays - 1)) {
       if (y_hat[j] < 0) {
          daily_pnl[j] = -RET[j+1,i]
       }
       else {
          daily_pnl[j] = RET[j+1,i];
       }
    }
    cum_pnl = cumsum(daily_pnl);
    CM = cbind(CM,cum_pnl);
    mean_pnl = mean(daily_pnl, na.rm=TRUE);
    yearly_pnl = mean(daily_pnl, na.rm=TRUE) * 252;
    total_pnl = sum(daily_pnl, na.rm = TRUE);
    sharpe = compute_Sharpe_Ratio(daily_pnl);

    stats_this_stock = c(sharpe, mean_pnl, yearly_pnl, total_pnl)
    STATS = rbind(STATS,stats_this_stock);

    readline('Press key to continue...');
  }

  rownames(STATS) = tickers;
  colnames(STATS) = c('sharpe', 'mean_pnl', 'yearly_pnl', 'total_pnl');;
  print(STATS);
  colnames(CM) = tickers;
  plot_cumsum(CM);

return('DONE!');
}
```

(c).

Stock Name	Sharpe	μ_{PNL}	Ret_{ann}	Ret_{total}
SPY	0.98	6.2×10^{-4}	0.156	0.779
VIX	0.83	3.9×10^{-3}	0.986	4.938
TNX	0.90	1.3×10^{-3}	0.326	1.633
OIL	1.26	1.5×10^{-3}	0.379	1.896
GOLD	0.92	1.3×10^{-3}	0.336	1.682
N225	6.18	5.0×10^{-3}	1.255	6.285
FTSE	2.61	1.6×10^{-3}	0.408	2.042
GLD	0.30	2.0×10^{-4}	0.052	0.258

(d). Here is the code for part d:

```
doWork_In_Sample = 1
if(doWork_In_Sample == 1){

   X_train = RET[1:nrDays - 1,];
   X_test = RET;
   STATS = NULL;
   CM = NULL;

   for (i in 1:nrTickers) {
      #print('##############################################');
      #print( paste('Stock = ',tickers[i] ) );

      daily_pnl = matrix(, nrow = nrDays - 101, ncol = 1);

      for (j in 101: (nrDays - 1)) {
         y_train = RET[ (j - 99) : j , tickers[i], drop=FALSE];
         X_train = RET[ (j - 100) : (j - 1) ,]
         y_hat = compute_linear_regression(X_train, y_train, X_test);
         if(y_hat[j] < 0) {
            daily_pnl[j - 100] = -RET[j + 1, i];
         }
         else {
            daily_pnl[j - 100] = RET[j + 1, i];
         }
      }

      cum_pnl = cumsum(daily_pnl);
      CM = cbind(CM,cum_pnl);
      mean_pnl = mean(daily_pnl, na.rm=TRUE);
      yearly_pnl = mean(daily_pnl, na.rm=TRUE) * 252;
      total_pnl = sum(daily_pnl, na.rm = TRUE);
      sharpe = compute_Sharpe_Ratio(daily_pnl);
      stats_this_stock = c(sharpe, mean_pnl, yearly_pnl, total_pnl)
      STATS = rbind(STATS,stats_this_stock);
   }

   rownames(STATS) = tickers;
   colnames(STATS) = c('sharpe', 'mean_pnl', 'yearly_pnl', 'total_pnl');
   print(STATS);
   colnames(CM) = tickers;
   plot_cumsum(CM);
}

return('DONE!');
}
```

(e).

Stock Name	Sharpe	μ_{PNL}	Ret_{ann}	Ret_{total}
SPY	0.26	1.6×10^{-4}	0.040	0.186
VIX	0.11	5.2×10^{-4}	0.131	0.605
TNX	0.55	8.1×10^{-4}	0.205	0.946
OIL	-0.40	-4.8×10^{-4}	-0.121	-0.556
GOLD	-0.25	-3.6×10^{-4}	-0.091	-0.418
N225	5.58	4.5×10^{-3}	1.145	5.279
FTSE	1.47	9.0×10^{-4}	0.227	1.048
GLD	0.15	1.0×10^{-4}	0.025	0.117

We have much lower values in (e) than (c), which denote a much less profitable prediction with our second model. This makes sense since we had much less data to work with our sliding window approach, but that method may be a better prediction model. The model in (c) may over-fit, which would result in the very high Sharpe ratios and returns.