# STAT 4130: Homework 1

Due: 2023-06-04

## Question 1

- For random variables, if their covariance is equal to zero, then they are uncorrelated.
- For random variables, their inner product is denoted by $< X, Y >= E(XY)$, if $E(XY) = 0$, they are called orthogonal, and denoted by $\perp$

1. Suppose we have a random variable X with a uniform distribution $X \sim \text{unif}(-1, 2)$, and $Y = h(X) = kX^2$, where is $(k \neq 0)$ is constant. Is X and Y correlated? independent? orthogonal?
2. Suppose we have a random variable X with a uniform distribution $X \sim \text{unif}(-1, 2)$, and $Y = h(X) = kX^2 + c$, where is $(k \neq 0, \quad c \neq 0)$ are both constants. Is X and Y correlated? independent? orthogonal?
3. Suppose we have a random variable X with a uniform distribution $X \sim \text{unif}(-1, 1)$, and $Y = h(X) = kX^2$, where is $(k \neq 0)$ is constant. Is X and Y correlated? independent? orthogonal?
4. If X and Y are uncorrelated, will X,Y be orthogonal? What is your conclusion about correlation and orthogonality?

## Question 2

Thus far, we only considered regression with scalar-valued outcomes. In some applications, the outcome is itself a vector: $\mathbf{y}_i \in \mathbf{R}^K$. We posit the relationship between the features and the vector-valued outcome is linear:

$$\mathbf{y}_i^T \approx \mathbf{x}_i^T \widehat{B}, \tag{1}$$

for some matrix of regression coefficients $\widehat{B} \in \mathbf{R}^{p \times K}$.

1. The sum of squared residuals (SSR) here is

$$\text{SSR}(B) \triangleq \sum_{i=1}^{n} \sum_{k=1}^{K} \frac{1}{2} (\mathbf{y}_{i,k} - \mathbf{x}_i^T b_k)^2,$$

   where $b_k \in \mathbf{R}^p$ is the $k$-th column of $B$. Express $\text{SSR}(B)$ in matrix notation (i.e. without using any explicit summations).

   **Hint:** work out how to express the SSR in terms of $B$,

$$\mathbf{X} = \begin{bmatrix} - & \mathbf{x}_1^T & - \\ & \vdots & \\ - & \mathbf{x}_n^T & - \end{bmatrix} \in \mathbf{R}^{n \times p}, \text{ and } \mathbf{y} = \begin{bmatrix} - & \mathbf{y}_1^T & - \\ & \vdots & \\ - & \mathbf{y}_n^T & - \end{bmatrix} \in \mathbf{R}^{n \times K}.$$

2. Find a closed-form expression for the matrix of regression coefficients that minimizes the SSR; i.e. find a (closed-form) expression for $\widehat{B} \in \arg\min_{B \in \mathbf{R}^{p \times K}} \text{SSR}(B)$.

3. Instead of minimizing the SSR, we break up the problem into $K$ separate regression problems with scalar-valued responses. That is, we fit $K$ linear models of the form

$$\mathbf{y}_{i,k} \approx \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_k,$$

where $\mathbf{y}_{i,k}$ is $k$-th outcome of the $i$-th sample (i.e. the $k$-th entry of $\mathbf{y}_i$) and $\widehat{\boldsymbol{\beta}}_k \in \mathbf{R}^p$ are the regression coefficients of the $k$-th linear model. How are the fitted coefficients from the $K$ separate regressions $\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_K$ related to the matrix of regression coefficients that minimizes the SSR $\widehat{B}$?

## Question 3

In this problem, we will predict the per capita crime using the other variables in the `Boston` dataset. The data can be imported as following:

```r
library(MASS)
data("Boston")
```

1. For each predictor, fit a simple linear regression model to predict the response. In which of the simple linear models is there a statistically significant association between the predictor and the response. (Note: there are 13 predictors, so please fit 13 models)
2. Fit a (multiple) regression model to predict the response using all the other features in the dataset. For which features can we reject the null $H_0 : \beta_j = 0$.
3. How do the results from (1) and (2) compare. Create a scatterplot displaying the simple regression coefficient of each predictor from (1) on the $x$-axis, and the multiple regression coefficient from (2) on the $y$-axis. That is, each predictor is displayed as a point on the plot.
4. Is there evidence of non-linear relationship between any of the features and response? For each predictor $\mathbf{x}_j$, look at the fit of the cubic model

$$\mathbf{y} \sim \beta_0 + \beta_1 \mathbf{x}_j + \beta_2 \mathbf{x}_j^2 + \beta_3 \mathbf{x}_j^3.$$