BST 140.751 final exam

Notes:

- You may not use a calculator for this exam.
- Please be neat and write legibly. Use the back of the pages if necessary.
- Good luck!

____

printed name

1. Let Y be an $n \times 1$ vector, X be an $n \times p$ full column rank matrix. Let $\mathcal{P} = \{\tilde{Y} \mid \tilde{Y} = X\beta \text{ for } \beta \in \mathbb{R}^p\}$. (I.e. $\mathcal{P}$ is the linear subspace spanned by the columns of X.)

   A. Derive a function $f : \mathbb{R}^n \to \mathbb{R}^n$ that takes any data point in $\mathbb{R}^n$ and finds its closest point in $\mathcal{P}$. (For Euclidean distance.) Call this point $\hat{Y}$.

   B. Argue that the difference $Y - \hat{Y}$ is orthogonal to any point in $\mathcal{P}$.

   C. Suppose that the distance is defined as $d(Y, \tilde{Y}) = (Y - \tilde{Y})'\Sigma^{-1}(Y - \tilde{Y})$ (Mahalanobis distance) where Σ is positive definite. Derive a function $g : \mathbb{R}^n \to \mathbb{R}^n$ that takes any data point in $\mathbb{R}^n$ and finds its closest point in $\mathcal{P}$ where the distance is the Mahalanobis distance.

2. Consider the model $Y|\beta \sim N(X\beta, \Sigma)$ where Y is a vector of length N, β is a vector of length p and X is full column rank. Suppose that we assume $\beta \sim N(\mu, \Gamma)$. What is the distribution of $\beta|Y$?

3. Let X be an $n \times p$ full rank design matrix that contains an intercept, Y be a vector of length n and $\mathbf{1}$ be a vector of length n of ones. Let $\bar{Y}$ be the sample average of the Y.

   A. Argue that $X(X'X)^{-1}X'\mathbf{1} = \mathbf{1}$.

   B. Argue that $||Y - \bar{Y}\mathbf{1}||^2 = ||Y - \hat{Y}||^2 + ||\hat{Y} - \bar{Y}\mathbf{1}||^2$.

   C. Assume that $Y \sim N(X\beta, \sigma^2\mathbf{I})$. Calculate $E[||Y - \hat{Y}||^2]$, $E[||\hat{Y} - \bar{Y}\mathbf{1}||^2]$ and $E[||Y - \bar{Y}\mathbf{1}||^2]$.