CS 194-10, Fall 2011
Assignment 1 Solution

1. (15 pts) *Uncertainty of predictions made by linear regression*: The derivation goes through just as for the expected value, except a bit more complicated. First, we note that

$$\mathbf{Y} - E[\mathbf{Y}] = \mathbf{X}\mathbf{w} + \epsilon - E[\mathbf{X}\mathbf{w} + \epsilon] = \mathbf{X}\mathbf{w} + \epsilon - E[\mathbf{X}\mathbf{w}] = \epsilon$$

   and then

$$\begin{aligned}
Var[\hat{\mathbf{w}}] &= E[(\hat{\mathbf{w}} - E(\hat{\mathbf{w}}))(\hat{\mathbf{w}} - E(\hat{\mathbf{w}}))^T] \\
&= E[((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - E((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}))((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - E((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}))^T] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E[(\mathbf{Y} - E(\mathbf{Y}))((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - E((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}))^T] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E[(\mathbf{Y} - E(\mathbf{Y}))(\mathbf{Y} - E(\mathbf{Y}))^T]((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E[\epsilon\epsilon^T]((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\sigma^2\mathbf{I})((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T\sigma^2 \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}((\mathbf{X}^T\mathbf{X})^{-1})^T\sigma^2 \\
&= (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1}\sigma^2 \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2\,.
\end{aligned}$$

2. (15 pts) *Weighted regression*:

   (a) As suggested in the question, we rewrite as follows:

$$\begin{aligned}
\hat{\mathbf{w}} &= \arg\min_{\mathbf{w}} \sum_{i=1}^{N} F_i(y_i - \mathbf{w}^T\mathbf{x}_i)^2 \\
&= \arg\min_{\mathbf{w}} \sum_{i=1}^{N} (\sqrt{F_i}y_i - \mathbf{w}^T\sqrt{F_i}\mathbf{x}_i)^2 \\
&= \arg\min_{\mathbf{w}} \sum_{i=1}^{N} (G_i y_i - \mathbf{w}^T G_i\mathbf{x}_i)^2
\end{aligned}$$

   where $G_i = \sqrt{F_i}$.

   (b) Now put the G_is as diagonal elements of a matrix $\mathbf{G}$, with zeroes elsewhere. Then the modified matrices are $\mathbf{X}' = \mathbf{G}\mathbf{X}$ and $\mathbf{Y}' = \mathbf{G}\mathbf{Y}$.

   (c) The normal equations are

$$\begin{aligned}
\hat{\mathbf{w}} &= ((\mathbf{X}')^T\mathbf{X}')^{-1}(\mathbf{X}')^T\mathbf{Y}' \\
&= ((\mathbf{G}\mathbf{X})^T\mathbf{G}\mathbf{X})^{-1}(\mathbf{G}\mathbf{X})^T\mathbf{G}\mathbf{Y} \\
&= (\mathbf{X}^T\mathbf{G}^T\mathbf{G}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{G}^T\mathbf{G}\mathbf{Y} \\
&= (\mathbf{X}^T\mathbf{F}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{F}\mathbf{Y}
\end{aligned}$$

   where $\mathbf{F}$ is the matrix with F_is along the diagonal, and we have used $\mathbf{G}^T\mathbf{G} = \mathbf{G}\mathbf{G} = \mathbf{F}$.