Franklin Hu, Sunil Pedapudi
SID: 20157715, SID: 20247144
CS 194-10
2011-09-09
Assignment 1

1. Uncertainty of predictions made by linear regression
   We first note the following:
   Given
$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^{\mathrm{T}}],$$
   for $X = \epsilon$,
$$Var(\epsilon) = \mathbb{E}[(\epsilon - \mathbb{E}[\epsilon])(\epsilon - \mathbb{E}[\epsilon])^{\mathrm{T}}] = \mathbb{E}[(\epsilon - 0)(\epsilon - 0)^{\mathrm{T}}] = \mathbb{E}[\epsilon\epsilon^{\mathrm{T}}]$$
   Then,
$$\begin{aligned}
Var(\hat{w}) &= \mathbb{E}[(\hat{w} - \mathbb{E}[\hat{w}])(\hat{w} - \mathbb{E}[\hat{w}])^{\mathrm{T}}] \\
&= \mathbb{E}[(\hat{w} - \mathbb{E}[\hat{w}])(\hat{w}^{\mathrm{T}} - \mathbb{E}[\hat{w}]^{\mathrm{T}})] \\
&= \mathbb{E}[\hat{w}\hat{w}^{\mathrm{T}} - \hat{w}\mathbb{E}[\hat{w}]^{\mathrm{T}} - \mathbb{E}[\hat{w}]\hat{w}^{\mathrm{T}} + \mathbb{E}[\hat{w}]\mathbb{E}[\hat{w}]^{\mathrm{T}}] \\
&= \mathbb{E}[\hat{w}\hat{w}^{\mathrm{T}} - \hat{w}w^{\mathrm{T}} - w\hat{w}^{\mathrm{T}} + w\hat{w}^{\mathrm{T}}] \\
&= \mathbb{E}[\hat{w}\hat{w}^{\mathrm{T}}] - \mathbb{E}[\hat{w}]w^{\mathrm{T}} - w\mathbb{E}[\hat{w}^{\mathrm{T}}] + ww^{\mathrm{T}} \\
&= \mathbb{E}[\hat{w}\hat{w}^{\mathrm{T}}] - ww^{\mathrm{T}} - ww^{\mathrm{T}} + ww^{\mathrm{T}} \\
&= \mathbb{E}[(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y((X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y)^{\mathrm{T}}] - ww^{\mathrm{T}} \\
&= \mathbb{E}[(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}YY^{\mathrm{T}}X(X^{\mathrm{T}}X)^{-1}] - ww^{\mathrm{T}} \\
&= \mathbb{E}[(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(Xw + \epsilon)(Xw + \epsilon)^{\mathrm{T}}X(X^{\mathrm{T}}X)^{-1}] - ww^{\mathrm{T}} \\
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\mathbb{E}[Xww^{\mathrm{T}}X^{\mathrm{T}} + Xw\epsilon + \epsilon w^{\mathrm{T}}X^{T} + \epsilon\epsilon^{\mathrm{T}}]X(X^{\mathrm{T}}X)^{-1} - ww^{\mathrm{T}} \\
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(\mathbb{E}[Xww^{\mathrm{T}}X^{\mathrm{T}}] + Xw\mathbb{E}[\epsilon] + \mathbb{E}[\epsilon]w^{\mathrm{T}}X^{\mathrm{T}} + \mathbb{E}[\epsilon\epsilon^{\mathrm{T}}])X(X^{\mathrm{T}}X) - ww^{\mathrm{T}} \\
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(\mathbb{E}[Xww^{\mathrm{T}}X^{\mathrm{T}}] + 0 + 0 + \sigma^2)X(X^{\mathrm{T}}X) - ww^{\mathrm{T}} \\
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Xww^{\mathrm{T}}X^{\mathrm{T}}X(X^{\mathrm{T}}X)^{-1} + (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\sigma^2X(X^{\mathrm{T}}X)^{-1} - ww^{\mathrm{T}} \\
&= ww^{\mathrm{T}} + (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}X(X^{\mathrm{T}}X)^{-1}\sigma^2 - ww^{\mathrm{T}} \\
&= (X^{\mathrm{T}}X)^{-1}\sigma^2
\end{aligned}$$

2. Weighted regression
   (a) Let $G_i = \sqrt{F_i}$. Then, we can reorganize
$$\begin{aligned}
F_i(y_i - \mathbf{w^Tx_i})^2 &= G_i^2(y_i - \mathbf{w^Tx_i})^2 \\
&= G_i^2y_i^2 - 2G_i^2y_i\mathbf{w^Tx_i} + G_i^2x_i^2 \\
&= (G_iy_i - G_i\mathbf{w^Tx_i})^2
\end{aligned}$$
   (b) Let us first express G_i as a matrix $\mathbf{G}$. Given
$$\mathbf{g} = \begin{pmatrix} G_1 \\ G_2 \\ \vdots \\ G_N \end{pmatrix}$$
   we can express $\mathbf{G} = \mathbf{g^TI}$. Then, we realize that
$$\mathbf{Y'} = \begin{pmatrix} G_1y_1 \\ G_2y_2 \\ \vdots \\ G_Ny_N \end{pmatrix}$$

$$\mathbf{X}' = \begin{pmatrix} G_1\mathbf{x_1} & G_2\mathbf{x_2} & \dots G_N & \mathbf{x_N} \end{pmatrix}$$

which is simply $\mathbf{X}' = \mathbf{GX}$ and $\mathbf{Y}' = \mathbf{GY}$.

(c) Since we have obtained a data matrix $\mathbf{X}'$ and a label matrix $\mathbf{Y}'$, we can substitute to find $\hat{w}$:

$$\hat{w} = (\mathbf{X'^T}\mathbf{X}')^{-1}\mathbf{X^T}\mathbf{Y} = ((\mathbf{GX})^T\mathbf{X})^{-1}(\mathbf{GX})^T\mathbf{GY}$$

3. Predicting travel times for seismic waves

   (a) Top P and S stations

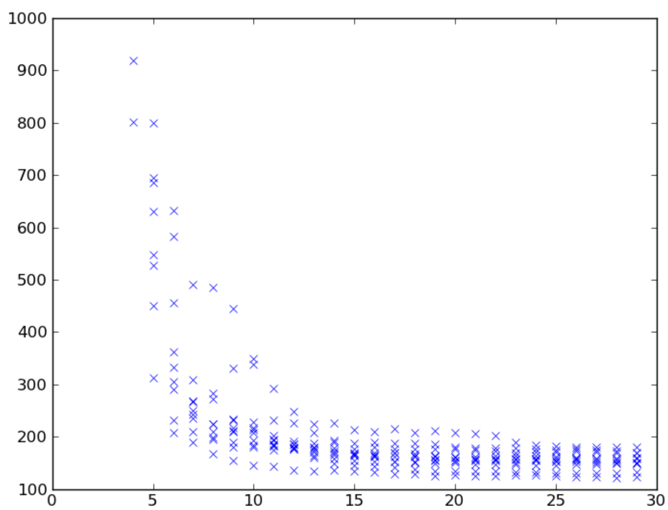
   - P_1 is station 1069 with 4239 events
   - P_2 is station 908 with 3996 events
   - S_1 is station 1069 with 486 events
   - S_2 is station 908 with 444 events

   (b) Time residuals at (0,0) with k=6 for top stations

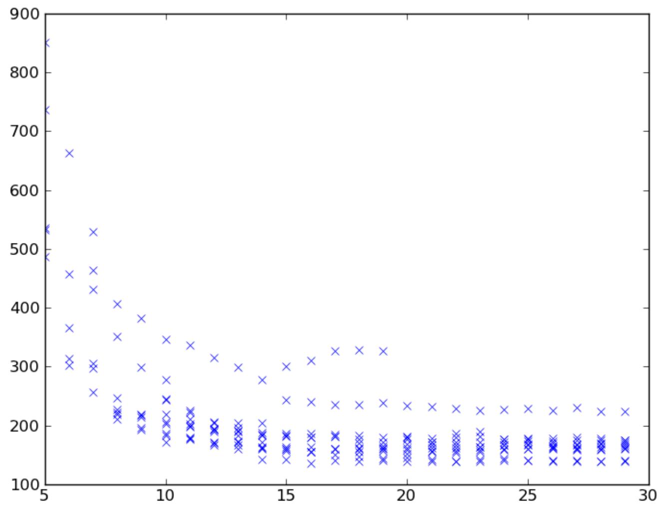
   - Station 1069, P: -0.17753345136616999
   - Station 908, P: -1.5740069952010929
   - Station 1069, S: 2.1081519933440749
   - Station 908, S: -30.214221599799373

   (c) Best k values

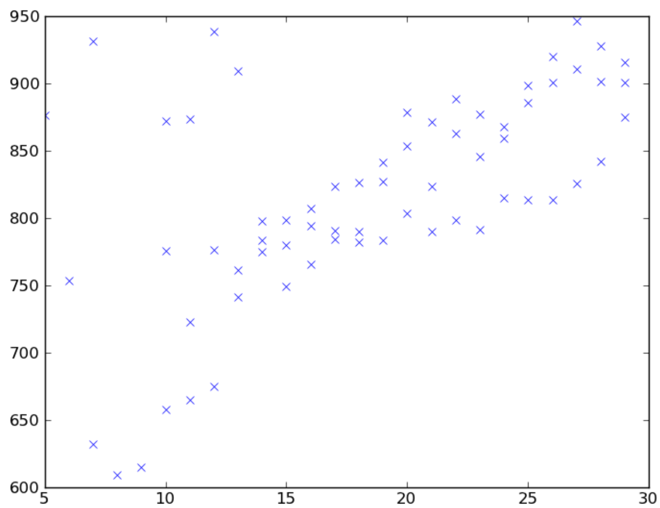
   - Station 1069, P: k=12



- Station 908, P: k=15



- Station 1069, S: k=8



- Station 908, S: k=58