Exercise 9

13.1

记

$$X = \begin{pmatrix} 1 & x_1 \\ 2 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$Y = (y_1, y_2, \cdots, y_n)^T$$

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$$

最小二乘法的结果为

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

其中回归结果的残差为

$$\begin{aligned} \hat{\varepsilon} &= Y - X\hat{\beta} \\ &= Y - X(X^T X)^{-1} X^T Y \\ &= (I - X(X^T X)^{-1} X^T) Y \\ &= (I - X(X^T X)^{-1} X^T)(X\beta + \varepsilon) \\ &= (I - X(X^T X)^{-1} X^T)\varepsilon \end{aligned}$$

残差平方和为

$$\begin{aligned} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 &= \hat{\varepsilon}^T \hat{\varepsilon} \\ &= \varepsilon^T (I - X(X^T X)^{-1} X^T)^T (I - X(X^T X)^{-1} X^T)\varepsilon \\ &= \varepsilon^T (I - X(X^T X)^{-1} X^T)\varepsilon \end{aligned}$$

求期望，注意到ε_i之间的独立性

$$\begin{aligned} E\sum_{i=1}^{n} \hat{\varepsilon}_i^2 &= E\varepsilon^T (I - X(X^T X)^{-1} X^T)\varepsilon \\ nE\hat{\varepsilon}_i^2 &= Etr(\varepsilon^T (I - X(X^T X)^{-1} X^T)\varepsilon) \\ &= tr(E(\varepsilon\varepsilon^T)(I - X(X^T X)^{-1} X^T)) \\ &= n\sigma^2 tr(I - X(X^T X)^{-1} X^T) \\ &= n\sigma^2 (n - tr(X^T X(X^T X)^{-1})) \\ &= n\sigma^2 (n-2) \end{aligned}$$

也就是

$$\sigma^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \frac{1}{n-2}\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

13.2

最小二乘法的结果为

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

其中

$$Y = X\beta + \varepsilon$$

于是

$$\mathbb{E}(\hat{\beta}|X^n) = \beta = \begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix}$$

下面计算方差

$$\begin{aligned}\hat{\beta}\hat{\beta}^T &= (X^TX)^{-1}X^TY\left[(X^TX)^{-1}X^TY\right]^T \\ &= (X^TX)^{-1}X^TYY^TX(X^TX)^{-1}\end{aligned}$$

其中

$$\mathbb{E}(YY^T) = \mathbb{E}(X\beta+\varepsilon)(X\beta+\varepsilon)^T = \mathbb{E}(X\beta)(X\beta)^T + \mathbb{E}\varepsilon\varepsilon^T = X\beta\beta^TX^T + \sigma^2 I$$

于是

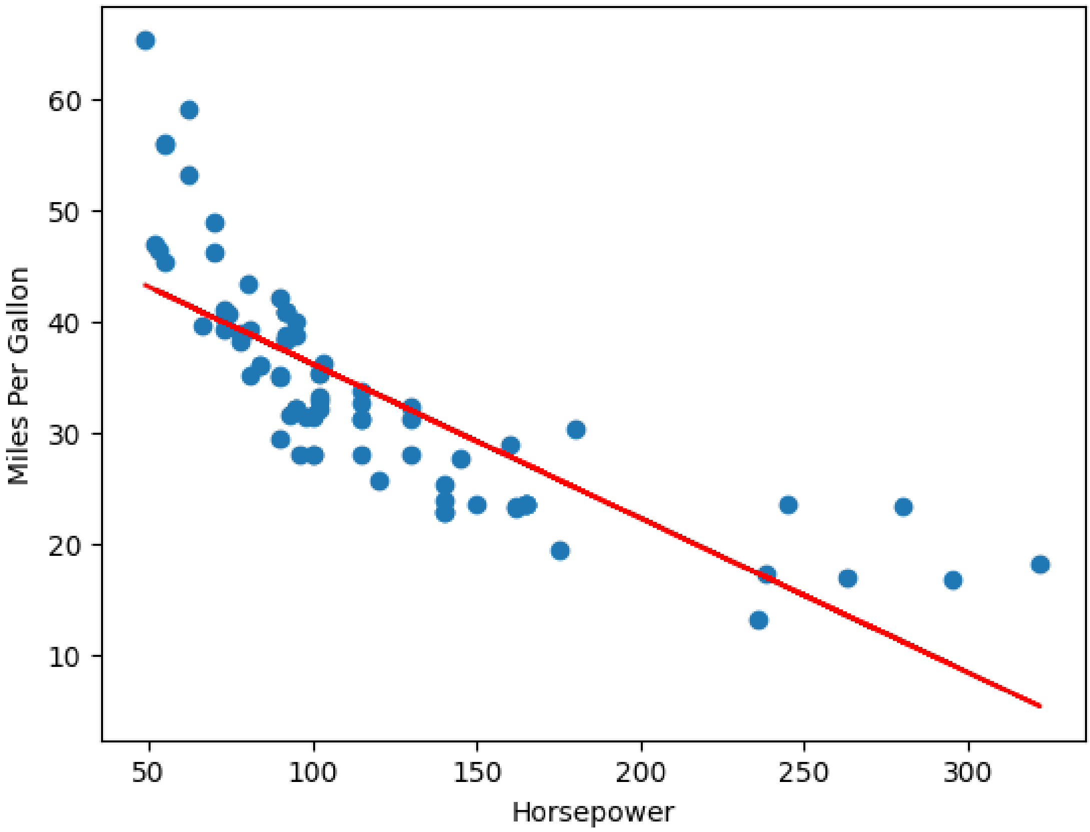
$$\begin{aligned}\mathbb{E}\left(\hat{\beta}\hat{\beta}^T\right) &= (X^TX)^{-1}X^T(X\beta\beta^TX^T + \sigma^2 I)X(X^TX)^{-1} \\ &= \beta\beta^T + \sigma^2(X^TX)^{-1}\end{aligned}$$

$$\begin{aligned}\mathbb{V}(\hat{\beta}|X^n) &= \mathbb{E}\left(\hat{\beta}\hat{\beta}^T\right) - \mathbb{E}\hat{\beta}\mathbb{E}\hat{\beta}^T \\ &= \sigma^2(X^TX)^{-1} \\ &= \frac{\sigma^2}{ns_X^2}\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n x_i^2 & -\overline{X}_n \\ -\overline{X}_n & 1\end{pmatrix}\end{aligned}$$

13.6

(a)

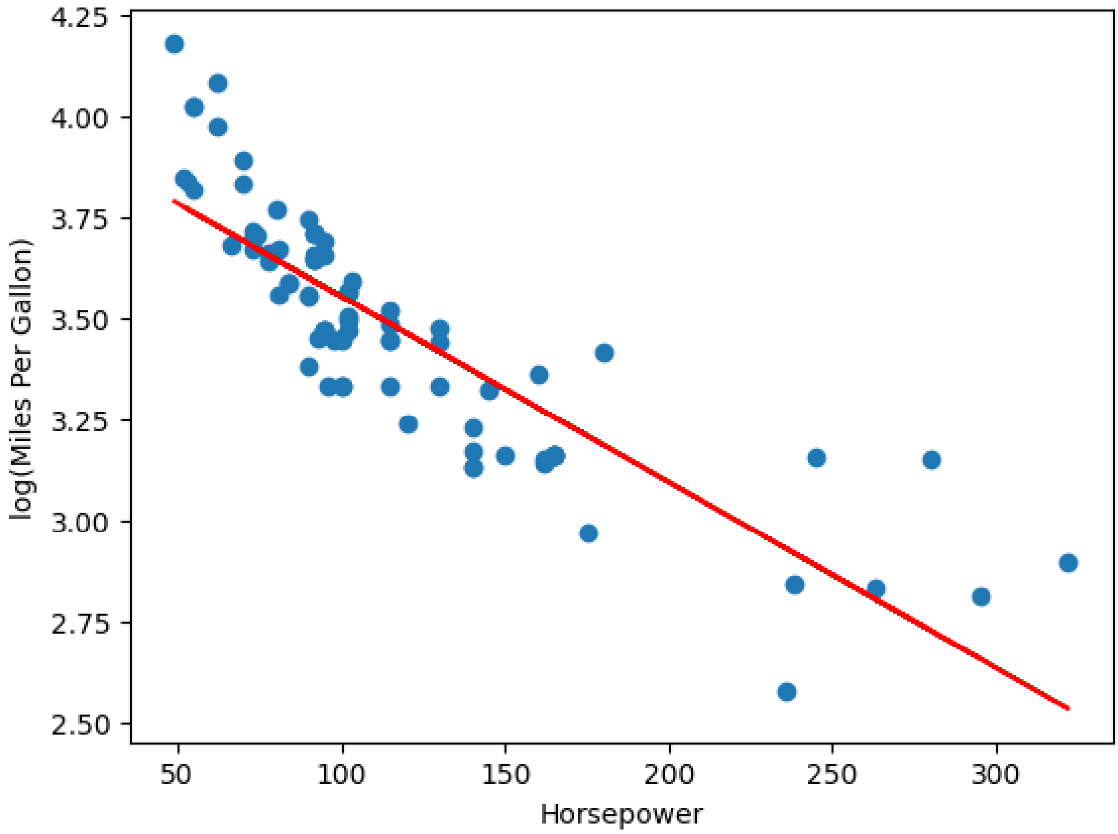
回归结果如为 MPG $= -0.13902326\,$HP $+ 50.06607807$.



此时相关系数为-0.79，线性回归结果在两端有较大偏差。

(b)

回归结果如为 $\log \mathrm{MPG} = -0.0045889\,\mathrm{HP} + 4.0132294$.



此时相关系数为-0.85，可见取对数后相关系数更接近-1，拟合效果更好。

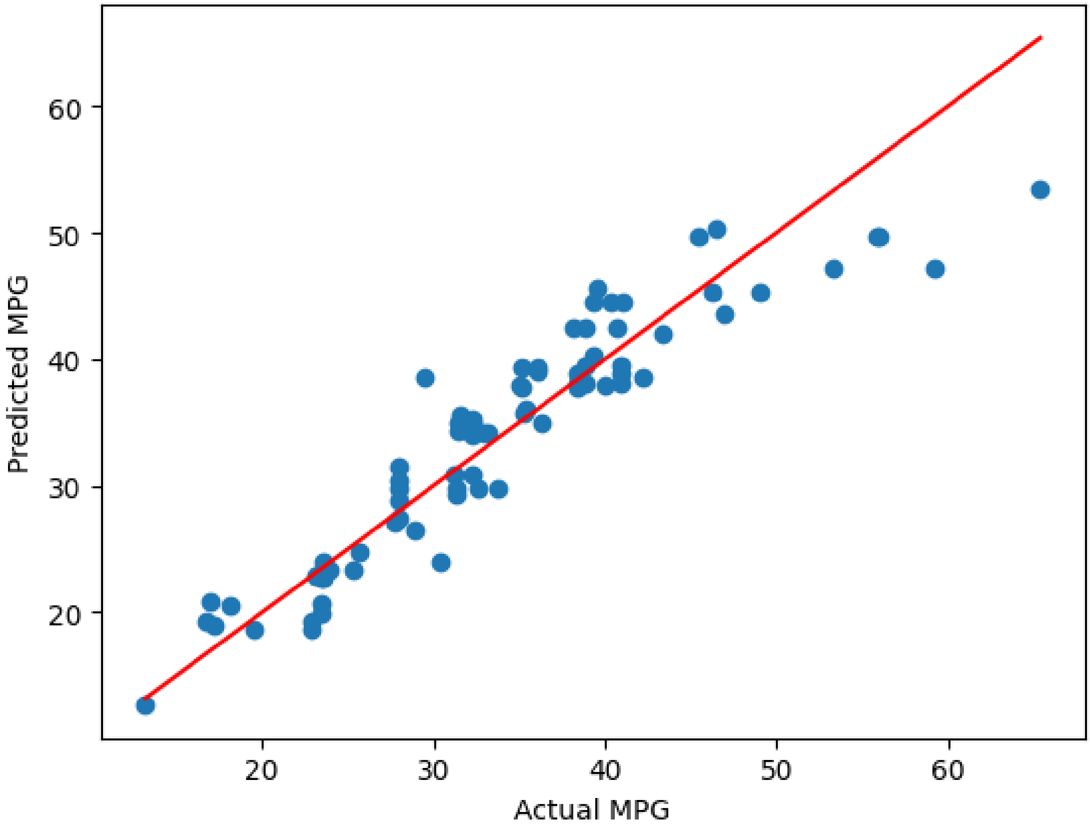
13.7

回归结果如下：

$$\mathrm{MPG} = 192.437753 - 0.0156450113\mathrm{HP} + 0.392212315\mathrm{SP} - 1.29481848\mathrm{WT} - 1.85980373\mathrm{VOL}$$

对比上一问中仅使用HP的回归结果，其prediction risk为3049；使用多维回归结果的prediction risk为1027，可见我们获得了更好的拟合效果。

下图展示了所有数据点真实值与预测值的对比：



4.2

(a)

$$
\begin{aligned}
f(x,y) &= \frac{1}{2\pi\sigma_{11}\sigma_{22}\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[(\frac{x-\mu_1}{\sigma_{11}})^2 - 2\rho(\frac{x-\mu_1}{\sigma_{11}})(\frac{y-\mu_2}{\sigma_{22}}) + (\frac{y-\mu_2}{\sigma_{22}})^2\right]} \\
&= \frac{1}{2\sqrt{3}\pi} e^{-\frac{2}{3}\left[\left(\frac{x}{2}\right)^2 - \left(\frac{x}{2}\right)(y-2) + (y-2)^2\right]}
\end{aligned}
$$

(b)

$$\Sigma_{12} = \Sigma_{21} = \rho\sigma_{11}\sigma_{22} = 1$$

$$\Sigma = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}$$

所以 squared generalized distance 为

$$\frac{{x_1}^2}{3} - \frac{2\,x_1\,x_2}{3} + \frac{4\,x_1}{3} + \frac{4\,{x_2}^2}{3} - \frac{16\,x_2}{3} + \frac{16}{3}$$

4.5

(a)

$$
\begin{aligned}
\mu_{1|x_2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) = x_2 - 2 \\
\Sigma_{11|x_2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = 3
\end{aligned}
$$

于是$X_1|x_2 \sim N(x_2 - 2, 3)$

$$f(X_1|x_2) = \frac{1}{\sqrt{6\pi}} e^{-\frac{1}{6}(X_1 - x_2 + 2)^2}$$

(b)

$$\mu_{2|x_1,x_3} = 1 - 2(x_1+3) = -2x_1 - 5$$
$$\Sigma_{22|x_1,x_3} = 5 - 4 = 1$$

于是$X_2|x_1,x_3 \sim N(-2x_1-5,1)$

$$f(X_2|x_1,x_3) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_2+2x_1+5)^2}$$

(c)

$$\mu_{1|x_2,x_3} = 0.5\,x_1 + 0.5\,x_2 + 1.5$$
$$\Sigma_{33|x_1,x_2} = 2 - 1.5 = 0.5$$

于是$X_3|x_1,x_2 \sim N(0.5\,x_1 + 0.5\,x_2 + 1.5, 0.5)$

$$f(X_3|x_1,x_2) = \frac{1}{\sqrt{\pi}} e^{-(X_3-0.5\,x_1-0.5\,x_2-1.5)^2}$$

4.16

(a) 由于X_1,X_2,X_3,X_4相互独立，所以

$$V_1 = \frac{1}{4}X_1 - \frac{1}{4}X_2 + \frac{1}{4}X_3 - \frac{1}{4}X_4 \sim N_p(0,\frac{1}{4}\Sigma)$$
$$V_2 = \frac{1}{4}X_1 + \frac{1}{4}X_2 - \frac{1}{4}X_3 - \frac{1}{4}X_4 \sim N_p(0,\frac{1}{4}\Sigma)$$

(b)

V_1,V_2的协方差为

$$\begin{aligned} conv(V_1,V_2) &= \mathbb{E}\left[(V_1-\mathbb{E}V_1)(V_2-\mathbb{E}V_2)^T\right] \\ &= \mathbb{E}\left[V_1V_2^T\right] \\ &= \mathbb{E}\left[\left(\frac{1}{4}X_1 - \frac{1}{4}X_2 + \frac{1}{4}X_3 - \frac{1}{4}X_4\right)\left(\frac{1}{4}X_1 + \frac{1}{4}X_2 - \frac{1}{4}X_3 - \frac{1}{4}X_4\right)^T\right] \end{aligned}$$

其中

$$\mathbb{E}X_iX_i^T = \Sigma + \mu\mu^T$$
$$\mathbb{E}X_iX_j^T = \mu\mu^T (i \neq j)$$

于是

$$conv(V_1,V_2) = 0_{p\times p}$$

有

$$\mu_{V_1,V_2} = 0_{2p\times 1}$$
$$\Sigma_{V_1,V_2} = \begin{pmatrix} \frac{1}{4}\Sigma & 0 \\ 0 & \frac{1}{4}\Sigma \end{pmatrix}$$

所以

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim N_{2p}\left(0, \begin{pmatrix} \frac{1}{4}\Sigma & 0 \\ 0 & \frac{1}{4}\Sigma \end{pmatrix}\right)$$

$$f(V_1,V_2) = \frac{1}{(2\pi)^p\sqrt{|\Sigma|}} e^{-2V_1^T\Sigma^{-1}V_1 - 2V_2^T\Sigma^{-1}V_2}$$