Exercises Week 4

Michele De Vita

March 22, 2018

1. The advantage of OLS to absolute deviation is the derivability because the second is not derivable in some points

2. *Normal equation*: $\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'y$.
   OLS estimation: $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$
   The solution is unique since $\mathbf{X}$ is definite positive , linearly independent ($rk(\mathbf{X}) = \#rows$) then invertible.
   If $\mathbf{X}$ is not linearly independent we can't calculate the term $(\mathbf{X}'\mathbf{X})^{-1}$

3. *MLE* require assuming a distribution of the error because to fit into non-linear data distributions. If the errors are distributed normally then *MLE* converges to *OLS*

4. $H = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
   Symmetric property:

$$
\begin{aligned}
&H = H' \\
&\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = (\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')' \\
&\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}((\mathbf{X}'\mathbf{X})^{-1})'\mathbf{X}' \\
&\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}((\mathbf{X}'\mathbf{X})')^{-1}\mathbf{X}' \\
&\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'
\end{aligned}
$$

   Idempotent:

$$
\begin{aligned}
&H \cdot H = \\
&\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \cdot \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \\
&\mathbf{X}\underbrace{((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X})}_{1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = H
\end{aligned}
$$

5. With MLE of a linear model we have a variance $\sigma^2_{ML} = \dfrac{\hat{\varepsilon}'\hat{\varepsilon}}{n}$.
   Since $\mathbb{E}\left[\hat{\varepsilon}'\hat{\varepsilon}\right] = (n-p)\cdot\sigma^2$ we have $\mathbb{E}\left[\sigma^2_{ML}\right] = \dfrac{n-p}{n}\cdot\sigma^2$ which is a biased estimator. From $\mathbb{E}\left[\hat{\varepsilon}'\hat{\varepsilon}\right]$ we can easily find an unbiased estimator $\hat{\sigma} = \dfrac{1}{n-p}\varepsilon'\varepsilon$

6. $\mathbf{X}'\varepsilon = \mathbf{X}'(\mathbf{I}-\mathbf{H})\mathbf{y} = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{H}\mathbf{y} = \mathbf{X}'\mathbf{y} - \underbrace{\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}_{\mathbf{1}}\mathbf{X}'\mathbf{y} = \mathbf{0}$.
   With the last result and assuming that the model has intercept we can assert that: $\sum_i \mathbf{1}\hat{\varepsilon}_{\mathbf{i}} = \sum_i \varepsilon_{\mathbf{i}} = 0$

7. The coefficient of determination $R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = \frac{\sum_i (\hat{y}_i - \overline{y})^2}{\sum_i (y_i - \overline{y})^2}$ vary between $[0,1]$ and explain the goodness of the fit of the regression. If $R^2 = 1$ we have a perfect fit while with $R^2 = 0$ we have $\forall i \;\; \hat{y}_i = \overline{y}$ so it means that every point $\hat{y}$ is in the mean. $R^2 = 0$ doesn't necessarily mean that the response is unrelated with the variables because there could be a non-linear relationship

8. For model M2 we can say that $R^2_{M2} \geq R^2_{M1}$. Instead for R^3_{M3} we can't say anything except if the data are distributed logarithmically then $R^2_{M3} \geq R^2_{M1}$

9. The three condition for compare model by R^2 are:
   - Same response variable
   - Same number of parameters
   - Must include intercept

10. Zero mean: $\mathbb{E}\left[\varepsilon\right] = 0$
    Homoscedasticity: $Cov\left(\varepsilon\right) = I \cdot \sigma^2$
    No correlation: $\varepsilon \cdot \mathbf{X} = 0$
    Normality: $\varepsilon \quad N(0, \sigma^2 I)$

11. $\mathbb{E}\left[\hat{\beta}\right] = \mathbb{E}\left[(X'X)^{-1}X'y\right] = (X'X)^{-1}X'\,\mathbb{E}\left[y\right] = (X'X)^{-1}X'X\beta = \beta$ We must assume that $\mathbb{E}\left[\varepsilon\right] = 0$

12. $$\begin{aligned} Cov(\hat{\beta}) &= Cov((X'X)^{-1}X'y) \\ &= (X'X)^{-1}X'Cov(y)((X'X)^{-1}X')' \\ &= \sigma^2(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

    Assumptions: $Cov(y) = \mathbf{I}\sigma^2,\; rk(X) = k+1$

13. An optimal prediction of y is $\mathbf{x_0'}\hat{\beta}$.
    A prediction is optimal if minimize the prediction error

14. If $X \sim N$ then $(\mathbf{AX} + b) \sim N$

15. $\frac{(\hat{\beta} - \beta)'(X'X)(\hat{\beta} - \beta)}{\sigma^2} \sim \chi^2_p$

16. A needed assumption is that $\lim_{n\to\infty} \frac{1}{n} X_n' X_n = V : V$ positive definite.
    This condition is true when, for example, the vectors x_i are i.i.d. (independent and identically distributed)

17. $\mathbb{E}\left[\hat{\varepsilon}\right] = \mathbb{E}\left[y\right] - X(X'X)^{-1}X'\,\mathbb{E}\left[y\right] = X\beta - \underbrace{X(X'X)^{-1}X'}_{I}X\beta = 0$

    $Cov(\hat{\varepsilon}) = Cov((\mathbf{I} - \mathbf{H})\mathbf{y}) = (\mathbf{I} - \mathbf{H})\sigma^2\mathbf{I}(\mathbf{I} - \mathbf{H})' = \sigma^2(\mathbf{I} - \mathbf{H})$

18. The residuals cannot be used to evaluate the homoscedasticity because they are not homoscedastic or uncorrelated. To solve this problem in practice, we use the standardization:

$$r_i = \frac{\hat{\varepsilon}_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}\iota$$