Lecture 2

Linear Regression (Ch. 2)

Regression → characterizing dependence amongst data
"representing" one set of data as a function of another set of data

Linear regression: representing one set of data as a LINEAR function of another set of data)

Simple Linear Regression: one variable linearly dependent on another variable

y_t, $t=0,1,\dots,n,\dots$ Pressure readings

x_t, $t=0,1,\dots,n,\dots$ Temperature readings

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \qquad \varepsilon_t \text{ - error term}$$

Most analytically tractable results → if we assume that ε_t-s are independent, identically distributed R.V.-s (random variables)
i.i.d

$$\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

Wow, I can use LS approach to estimate params β!

$$\mathcal{L}_{\varepsilon_t}(\beta_0, \beta_1) = \left[\frac{1}{\sqrt{2\pi}\,\sigma_\varepsilon}\right]^N e^{-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{N}(y_t - \beta_0 - \beta_1 x_t)^2}$$

To maximize this likelihood, I must <u>minimize</u>

$$\sum_{t=1}^{N}(y_t - \beta_0 - \beta_1 x_t)^2 = S(\beta_0, \beta_1)$$ regarding β_0 & β_1

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\substack{\beta_0 = \hat\beta_0 \\ \beta_1 = \hat\beta_1}} = 0 \Rightarrow -2\sum_{t=1}^{N}(y_t - \hat\beta_0 - \hat\beta_1 x_t) = 0 \Rightarrow$$

$$N \cdot \hat\beta_0 + \left(\sum_{t=1}^{N} x_t\right)\hat\beta_1 = \sum_{t=1}^{N} y_t \quad (1$$

$$\left.\frac{\partial S}{\partial \beta_1}\right|_{\substack{\beta_0 = \hat\beta_0 \\ \beta_1 = \hat\beta_1}} = 0 \Rightarrow -2\sum_{t=1}^{N}(y_t - \hat\beta_0 - \hat\beta_1 x_t)x_t = 0$$

$$\Rightarrow \hat\beta_0 \sum_{t=1}^{N} x_t + \hat\beta_1 \sum_{t=1}^{N} x_t^2 = \sum_{t=1}^{N} x_t y_t \quad (2$$

$$(1) \Rightarrow \hat\beta_0 = \frac{1}{N}\sum_{t=1}^{N} y_t - \hat\beta_1 \frac{1}{N}\sum_{t=1}^{N} x_t = \bar y - \hat\beta_1 \bar x \qquad (*)$$

Sub. $(*)$ ~~[illegible]~~ into (2) and you get

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{N} x_t y_t - \frac{1}{N}\sum_{t=1}^{N} x_t \sum_{t=1}^{N} y_t}{\sum_{t=1}^{N} x_t^2 - \frac{1}{N}\left(\sum_{t=1}^{N} x_t\right)^2} = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - (\bar{x})^2}$$

$$= \frac{\sum_{t=1}^{N}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{N}(x_t - \bar{x})^2}$$

Substituting $\hat{\beta}_1$ into (*) gives

$$\hat{\beta}_0 = \bar{y} - \bar{x} \cdot \frac{\sum_{t=1}^{N} x_t y_t - \frac{1}{N}\sum x_t \sum y_t}{\sum_{t=1}^{N} x_t^2 - \frac{1}{N}\left(\sum (x_t)\right)^2}$$

Properties

$E[\hat{\beta}_0] = \beta_0$ $\qquad$ $E[\hat{\beta}_1] = \beta_1$

Algebra on p. 51-58 also gives

$$Var[\hat{\beta}_0] = \sigma_\varepsilon^2 \left(\frac{1}{N} + \frac{\bar{x}^2}{\sum_{t=1}^{N} x_t^2}\right)$$

$N \uparrow \Rightarrow$

$\sigma_{\hat{\beta}_0}, \sigma_{\hat{\beta}_1} \downarrow$

$$Var[\hat{\beta}_1] = \sigma_\varepsilon^2 \frac{1}{\sum_{t=1}^{N} x_t^2}$$

Multivariate Linear Regression

Similar to simple regression, except that we express one variable as a linear combination of multiple other variables

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \dots + \beta_n x_{t,n} + \varepsilon_t$$

ε_t is IID, $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$

It's equivalent to analyze

$$y_t = \beta_1 x_{t,1} + \beta_2 x_{t,2} + \dots + \beta_n x_{t,n} + \varepsilon_t$$

Because I can just set $x_{t,1} = 1$ & get that offset term.

Again, I need to minimize

$$\sum (y_t - \beta_1 x_{t,1} - \beta_2 x_{t,2} - \dots - \beta_n x_{t,n})^2 = S(\beta_1, \beta_2, \dots, \beta_n)$$

regarding $\beta_1, \beta_2, \dots, \beta_n$

$$\left.\frac{\partial S(\beta_1, \beta_2, \dots, \beta_n)}{\partial \beta_1}\right|_{\vec{\beta} = \hat{\vec{\beta}}} = 0 \; ; \quad \left.\frac{\partial S(\beta_1, \dots, \beta_n)}{\partial \beta_2}\right|_{\vec{\beta} = \hat{\vec{\beta}}} = 0 \quad \dots$$

$$\dots \quad \left.\frac{\partial S(\beta_1, \dots, \beta_n)}{\partial \beta_n}\right|_{\vec{\beta} = \hat{\vec{\beta}}} = 0$$

Easy way to do this is using matrix notation

$$
\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ - & - & - & - \\ x_{N,1} & x_{N,2} & \cdots & x_{N,n} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} =
$$

$$
Y = \underline{X} \cdot \vec{\beta} + \vec{\varepsilon}
$$

$$
S(\vec{\beta}) = (Y - \underline{X} \cdot \vec{\beta})^T (Y - \underline{X} \cdot \vec{\beta}) =
$$

$$
= Y^T Y - Y^T \cdot X \cdot \vec{\beta} - \vec{\beta}^T \underline{X}^T \cdot Y + \vec{\beta}^T \underline{X}^T \underline{X} \cdot \vec{\beta}
$$

$$
\frac{\partial \vec{S}}{\partial \vec{\beta}} \Bigg|_{\vec{\beta} = \hat{\vec{\beta}}} = \begin{bmatrix} \frac{\partial S}{\partial \beta_1} \\ \frac{\partial S}{\partial \beta_2} \\ \vdots \\ \frac{\partial S}{\partial \beta_n} \end{bmatrix}_{\vec{\beta} = \hat{\vec{\beta}}} =
$$

$$
= -2 \underline{X}^T Y + 2 \underline{X}^T \underline{X} \hat{\vec{\beta}} = 0
$$

$$
\underbrace{\underline{X}^T \underline{X}} \hat{\vec{\beta}} = \underline{X}^T \underline{Y}
$$

If $\underline{X}$ is a full column rank matrix (i.e. if $X^T X$ is invertible)

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

Properties

$E[\hat{\vec{\beta}}] = \vec{\beta}$ (easy to prove that when $X^T X$ is invertible, this is an unbiased estimate of β-s)

$Var(\hat{\vec{\beta}}) = Var[(X^T X)^{-1} X^T y] =$

$= (\underline{X}^T \underline{X})^{-1} \cdot \sigma_\varepsilon^2$ //

Again, more x-s (bigger N), $\|\underline{X}^T \underline{X}\|$ will grow & we should get less variance in $\vec{\beta}$ (unless I am getting highly correlated observations & then the above is NOT true).

Special case of an AR(1) model is "Random Walk"

$\phi_1 = 1 \quad \Rightarrow \quad X_t = X_{t-1} + a_t \qquad a_t$: NIID $\mathcal{N}(0, \sigma_a^2)$

Substituting successively, we can see that

$$X_t = \sum_{k=0}^{t} a_t$$

and that

$$\hat{X}_t(1) = X_t$$

(best prediction of the next sample is the previous one)

First model of the stock market

Bachelier's hypothesis - 1900

Also important because it's a limiting case between "stable" & "unstable" time-series (we'll talk more about it when we discuss time-series analysis).

Autoregressive Model of Order 2 - AR(2)

$$X_3 = \phi_1 X_2 + \phi_2 X_1 + a_3$$

$$X_4 = \phi_1 X_3 + \phi_2 X_2 + a_4 \qquad \Rightarrow$$

$$\vdots$$

$$X_N = \phi_1 X_{N-1} + \phi_2 X_{N-2} + a_N$$

$$\overset{\underline{Y}}{\begin{bmatrix} X_3 \\ X_4 \\ \vdots \\ X_N \end{bmatrix}} = \overset{\underline{X}}{\begin{bmatrix} X_2 & X_1 \\ X_3 & X_2 \\ \cdots & \cdots \\ X_{N-1} & X_{N-2} \end{bmatrix}} \overset{\vec{\beta}}{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}} + \overset{\vec{\varepsilon}}{\begin{bmatrix} a_3 \\ a_4 \\ \vdots \\ a_N \end{bmatrix}}$$

Now, this looks like a multivariate regression

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{Y}$$

$$\hat{\sigma}_a^2 = \frac{1}{N-2-2} \sum_{t=3}^{N} (X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2})$$