Regression as Conditional Expectation

recall:

linear regression model:

$$Y = \beta_0 \cdot 1 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon$$

$$\Longrightarrow \quad Y = \beta^T X + \varepsilon$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \quad X = \begin{pmatrix} 1 \\ X_1 \\ \vdots \\ X_p \end{pmatrix}$$

Conditional expectation:

see slide 62

$$E(Y \mid X = x)$$

$$= E(\beta^T X + \varepsilon \mid X = x)$$

$$= E(\beta^T x + \varepsilon \mid X = x)$$

$$= E(\underbrace{\beta^T x}_{\text{not random}} \mid X = x) + E(\varepsilon \mid X = x)$$

assume: X, ε stoch. independent

$$= \overbrace{\beta^T x}^{} + \underbrace{E(\varepsilon)}_{=0}$$

$$\Longrightarrow \boxed{E(Y \mid X = x) = \beta^T x}$$

$\Longrightarrow$ the regression function $m(x) = \beta^T x$ can be written as the conditional expectation of Y given $X = x$

data set: Credit_subset.csv

with $Y \in \{0, 1\}$

problem: linear regression requires that Y is a continuous variable

$\Rightarrow$ Could we derive

$$E(Y \mid X = x) = ?$$

$$Y \in \{0,1\} \Rightarrow Y \sim B(1,p)$$
$$\Rightarrow E(Y) = p$$
$$\Rightarrow E(Y \mid X = x) = p_x$$

now: different p for every x

idea:

$$E(Y \mid X = x) = F(\beta^T x)$$

where $F(u) = \frac{1}{1+e^{-u}}$

cdf of the standard logistic distribution

in R: plogis

also: any other cdf would be possible

(the main issue is that

$E(Y \mid X = x) \in [0,1]$)