What is the probability of (x,y) occurring?

$$P((x,y)) = \int_{\Omega} P((x,y)|\beta) \cdot P(\beta)\, d\beta \qquad \Omega \text{ (domain of } \beta\text{, possible values of } \beta)$$

Let us assume that ε_i are ϕ mean: $\mathbb{E}_{(x,y)}[\varepsilon] = 0$

Likelihood fxn, The likelihood of observing a set of parameters given observations

$$L(\theta|x) = P(x|\theta)$$

Error of probability:
Deterministic fxn $Y(x,\beta)$ with additive Gaussian noise

$$P(\varepsilon_1, \dots, \varepsilon_n|\beta) = \prod_{i=1}^{n} P(\varepsilon_i|\beta) = \prod_{i=1}^{n} P(\underset{\uparrow \text{dependent}}{Y_i} - \underset{\uparrow \text{independent}}{X^T\beta}|\beta)$$

$$Y_i = X_i^T\beta + \varepsilon_i \Rightarrow Y_i - X_i^T\beta = \varepsilon_i$$

How would you suspect the RV ε_i behaves? IT should be normally distributed.

$$Y_i = x_i^T\beta + 0 \text{ , where is gaussian distributed.}$$

$$Y_i - \underbrace{x_i^T\beta}_{\sum_{j=1}^{d} x_j\beta_j} = 0$$

If ε is gaussian and normally distributed, then $\varepsilon_i = Y_i - x_i^T\beta$ can be fitted by least squares

$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2) \qquad \text{mean} = 0,\ \text{variance} = \sigma^2$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(\varepsilon_i)^2}{2\sigma^2}$$

Cost/Error fxn

We then take the log likelihood and want to minimize the cost:

$$\underset{\substack{\uparrow \\ \text{minimize cost}}}{-}\underset{\substack{\uparrow \\ \text{log}}}{\log} \underset{\substack{\uparrow \\ \text{likelihood fxn}}}{P(\varepsilon_1 \cdots \varepsilon_n|\beta)} = \sum_{i=1}^{n} \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(y_i - x^T\beta)^2}{2\sigma^2}\right)\right]$$

$$= \sum_{i=1}^{n} \frac{(y_i - x^T\beta)^2}{2\sigma^2} + \frac{N}{2}\log 2\pi\sigma^2$$

Because of sq-root and sum.

We will continue we deriving a linear regression equation built on the MLE method.

We will estimate $\hat{\beta}_{ML}$ that will be used for prediction (iff x^TX is a non-singular matrix).

The final prediction (inference of our $\hat{\beta}$ estimator is):

$$\underset{\substack{\downarrow \\ \text{Estimate scalar} \\ \text{n-dim vector}}}{\hat{Y}_{i+1}} = x_{i+1}^T \underset{\substack{\downarrow \\ \text{MLE estimator}}}{\hat{\beta}_{ML}} = x_{i+1}^T \underbrace{(x^TX)^{-1} X^TY}_{\text{Comes from what you estimated}}$$