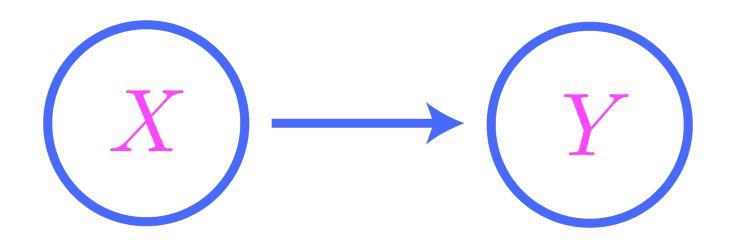
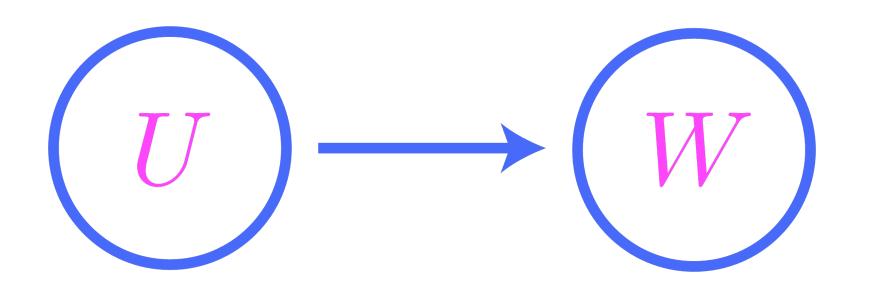
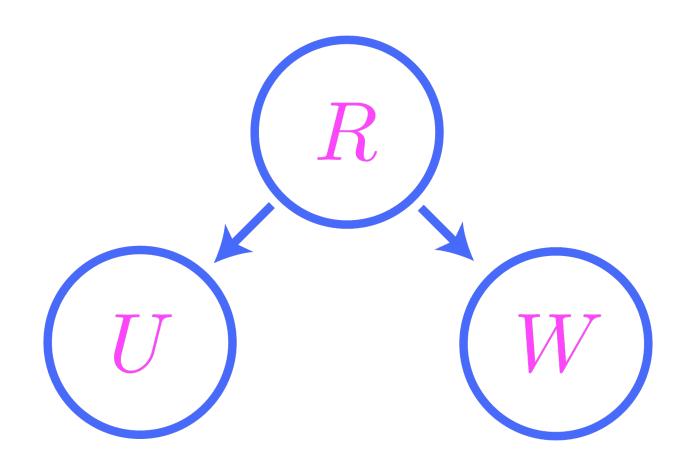
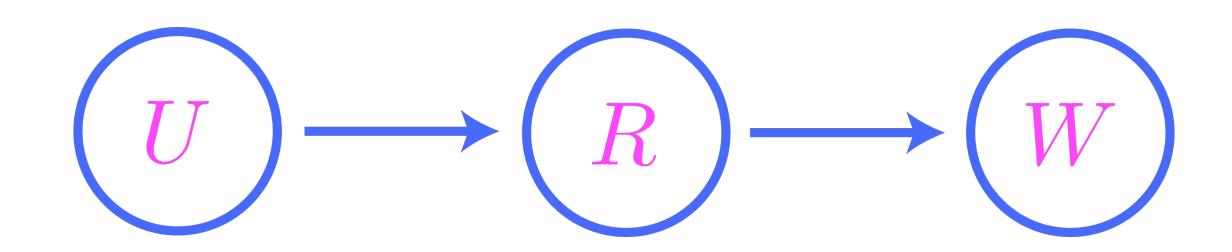
12. Probabilistic graphical models

12.1. A brief look at causal inference









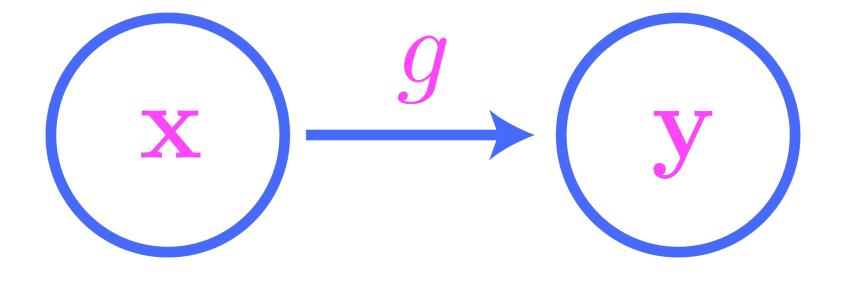
Causal structures and probability.

- Relationships of cause and effect represent strictly more structure than a joint probability distribution.
- A causal structure refines a joint probability distribution; it encodes more knowledge.
- The mapping from causal structures to joint probability distributions is many-to-one.

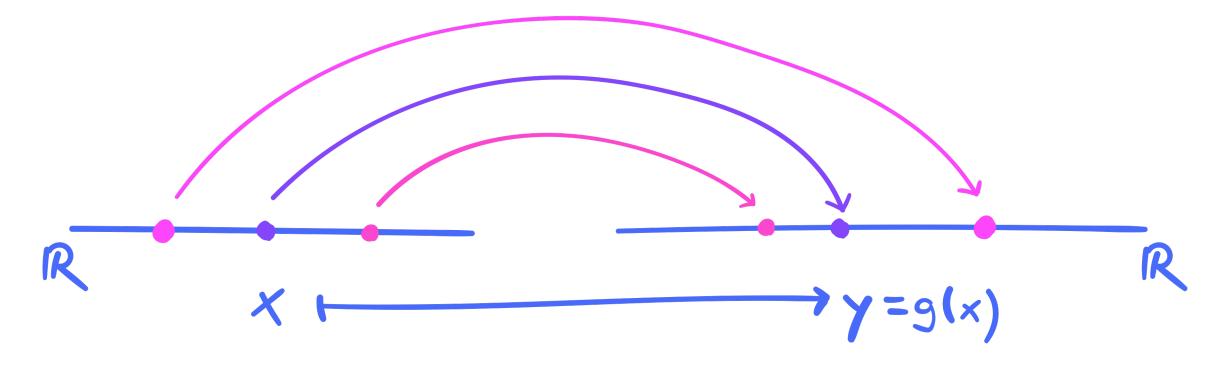


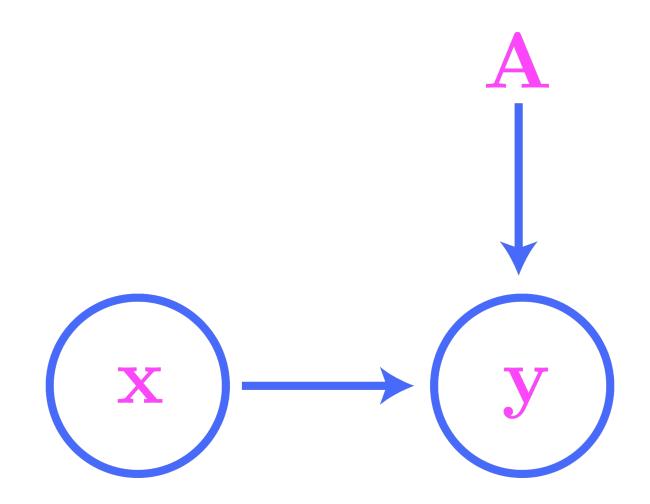
Do problem 1 on the worksheet.

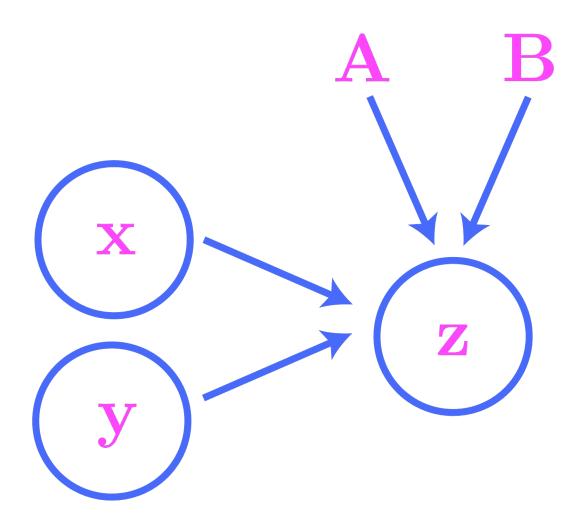
12.2. General probabilistic graphical models

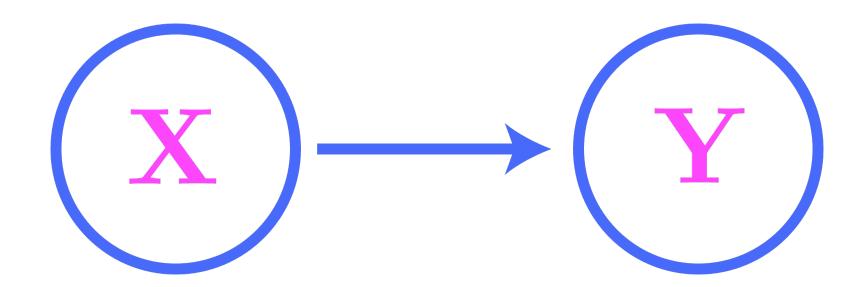


Deterministic link

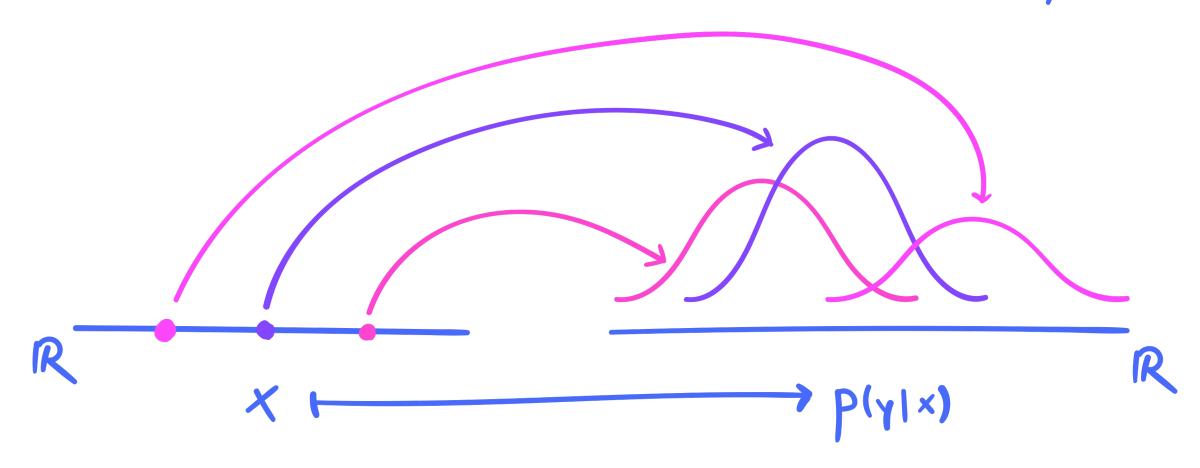






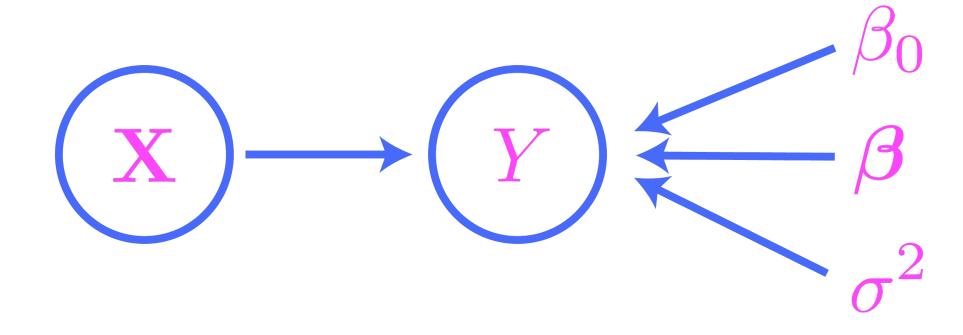


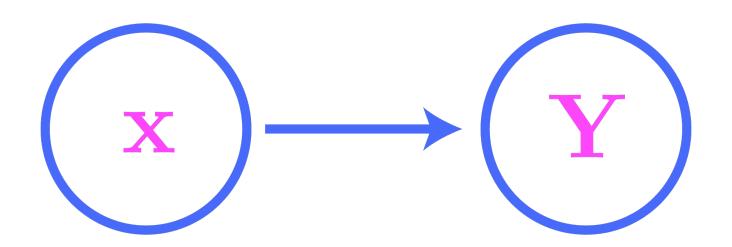
Stochastic link (Markov Kernel)

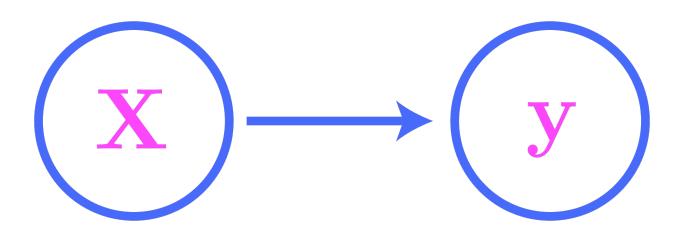


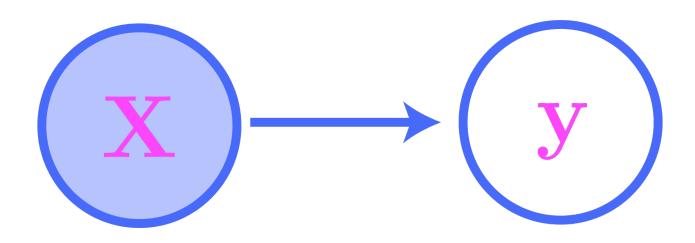


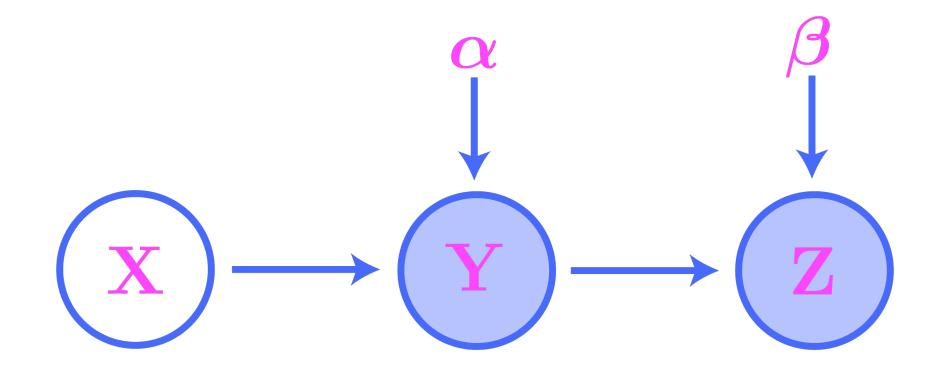
Do problem 2 on the worksheet.

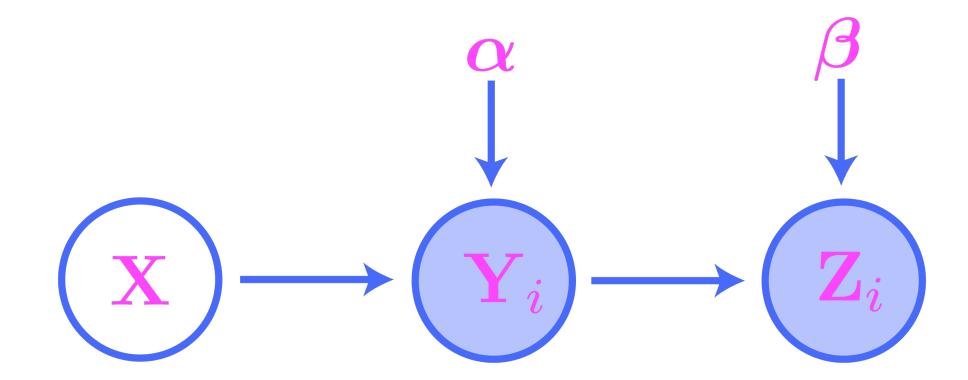


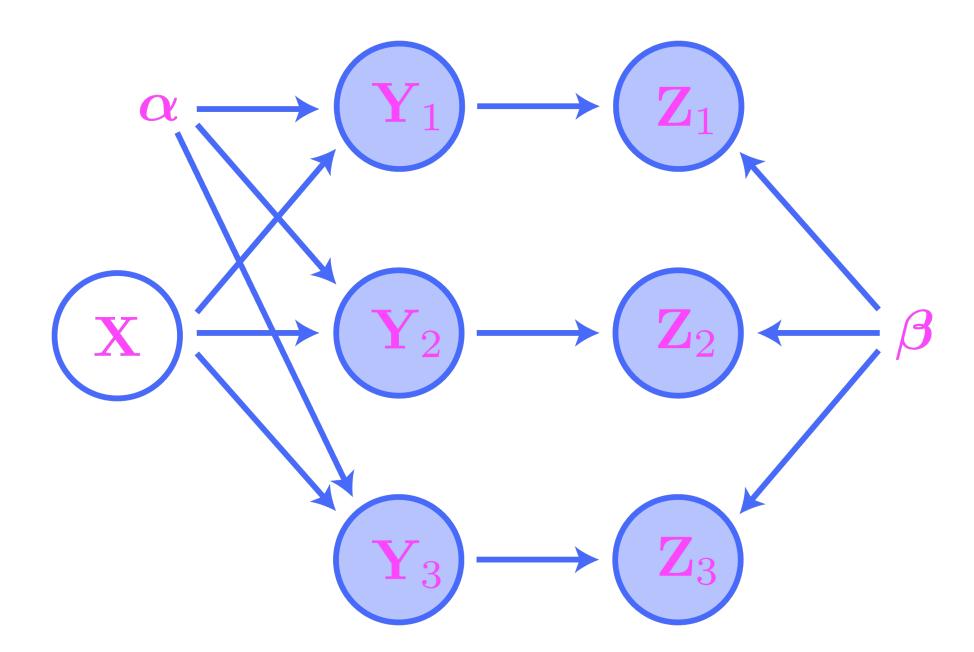


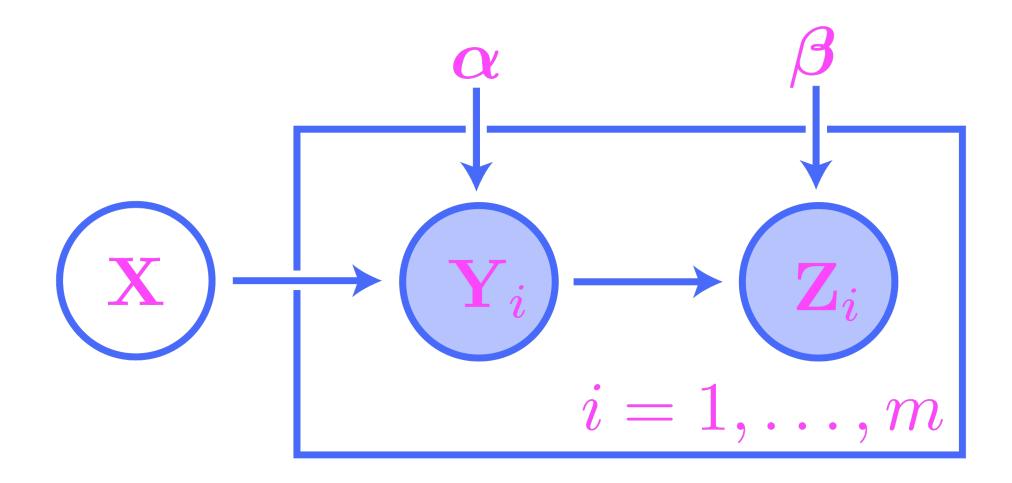














Do problem 3 on the worksheet.

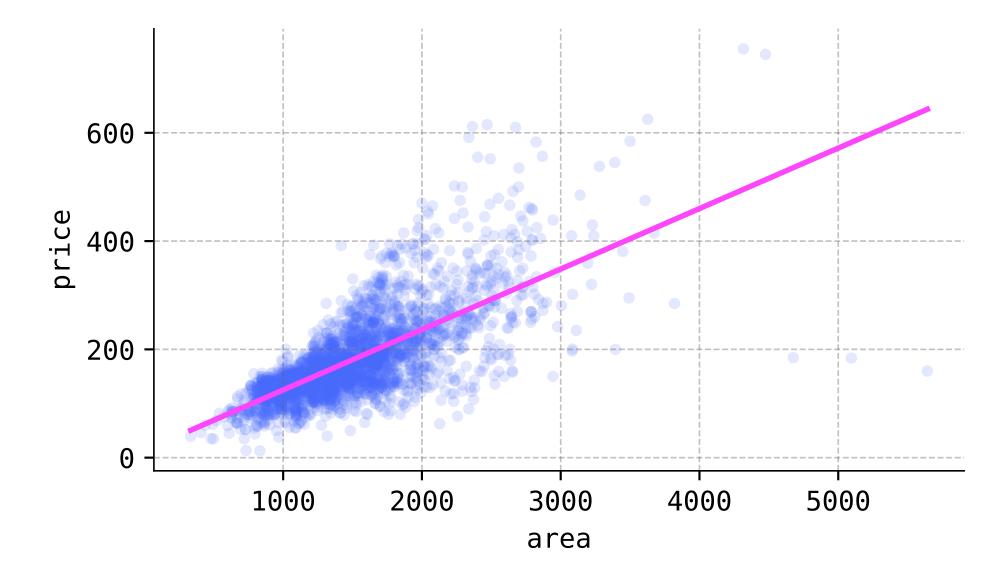


Definition 12.1

A probabilistic graphical model (PGM) consists of the following:

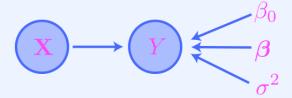
- 1. A set of vectors, some random and some deterministic, and some marked as observed and all others as hidden.
- 2. A graphical structure depicting the vectors as nodes and flows of influence (or information) as arrows between the nodes. If any of these flows are parametrized, then the graphical structure also has (un-circled) nodes for the parameters.
- 3. Mathematical descriptions of the flows as (possibly parametrized) link functions.

12.3. Linear regression models



Definition 12.2

A *linear regression model* is a probabilistic graphical model whose underlying graph is of the form



where $\mathbf{X} \in \mathbb{R}^n$. The model has the following parameters:

- A real parameter $\beta_0 \in \mathbb{R}$.
- A parameter vector $oldsymbol{eta} \in \mathbb{R}^n$.
- A positive real parameter $\sigma^2 > 0$.

The link function at Y is given by

$$Y \mid \mathbf{X} \sim \mathcal{N}ig(\mu, \sigma^2ig), \quad ext{where} \quad \mu = eta_0 + \mathbf{x}^\intercal oldsymbol{eta}.$$

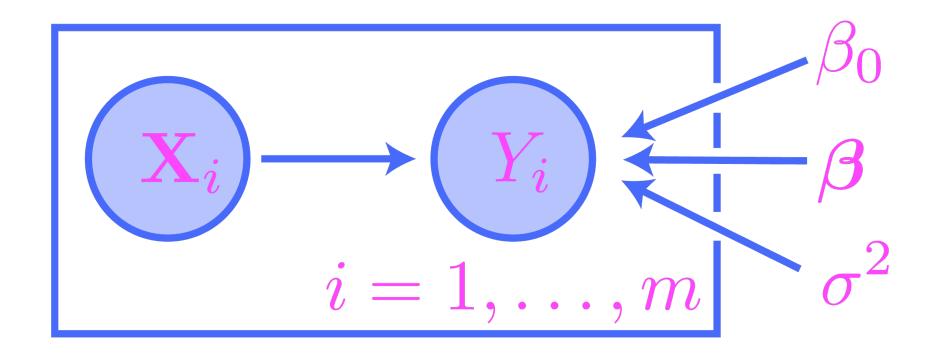
Å

Definition 12.3

For fixed $\mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$, the *model likelihood function* for a linear regression model is the function

$$\mathcal{L}(eta_0,oldsymbol{eta},\sigma^2;\ y\mid \mathbf{x}) \stackrel{ ext{def}}{=} fig(y\mid \mathbf{x};\ eta_0,oldsymbol{eta},\sigma^2ig) = rac{1}{\sqrt{2\pi\sigma^2}} ext{exp}\left[-rac{1}{2\sigma^2}(y-\mu)^2
ight] \quad ext{(12.2)}$$

of the parameters $eta_0, oldsymbol{eta}, \sigma^2$, where $\mu = eta_0 + \mathbf{x}^\intercal oldsymbol{eta}$.



Definition 12.4

Given an observed dataset

$$(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\ldots,(\mathbf{x}_m,y_m)\in\mathbb{R}^n imes\mathbb{R},$$

the data likelihood function for a linear regression model is the function

$$\mathcal{L}(eta_0,oldsymbol{eta},\sigma^2;y_1,\ldots,y_m\mid \mathbf{x}_1,\ldots,\mathbf{x}_m)\stackrel{ ext{def}}{=} fig(y_1,\ldots,y_m\mid \mathbf{x}_1,\ldots,\mathbf{x}_m;\ eta_0,oldsymbol{eta},\sigma^2ig)$$

of the parameters $\beta_0, \boldsymbol{\beta}, \sigma^2$.

Theorem 12.1 (Data likelihood functions of linear regression models)

Given an observed dataset

$$(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\ldots,(\mathbf{x}_m,y_m)\in\mathbb{R}^n imes\mathbb{R},$$

the data likelihood function for a linear regression model is given by

$$egin{aligned} \mathcal{L}_{ ext{data}}(eta_0,oldsymbol{eta},\sigma^2) &= \prod_{i=1}^m \mathcal{L}(eta_0,oldsymbol{eta},\sigma^2;\; y_i \mid \mathbf{x}_i) \ &= rac{1}{(2\pi\sigma^2)^{m/2}} ext{exp}\left[-rac{1}{2\sigma^2} \sum_{i=1}^m (y_i - \mu_i)^2
ight], \end{aligned}$$

where $\mu_i = eta_0 + \mathbf{x}_i^\intercal \boldsymbol{\beta}$ for each $i = 1, \dots, m$.



Do problems 3 and 4 on the worksheet.

