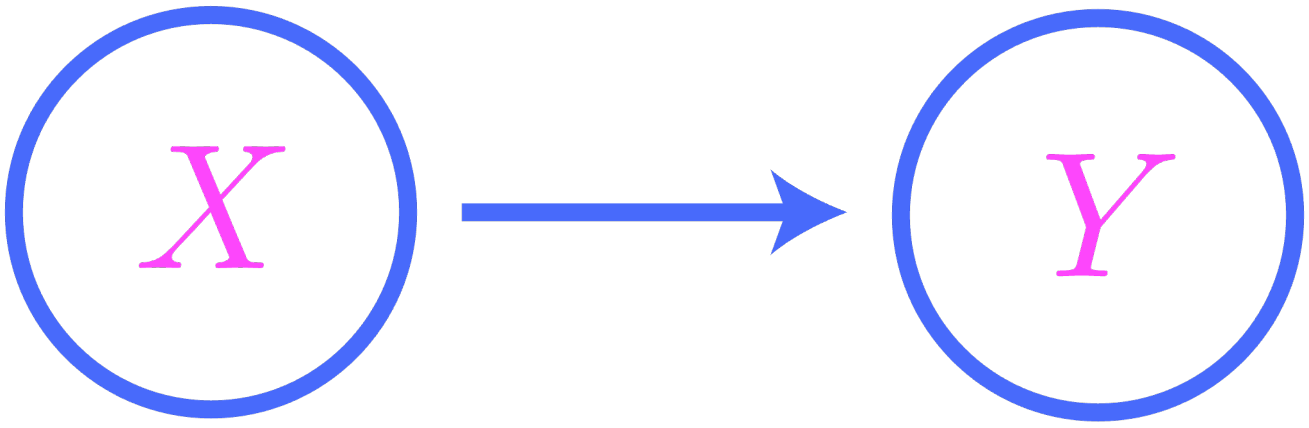
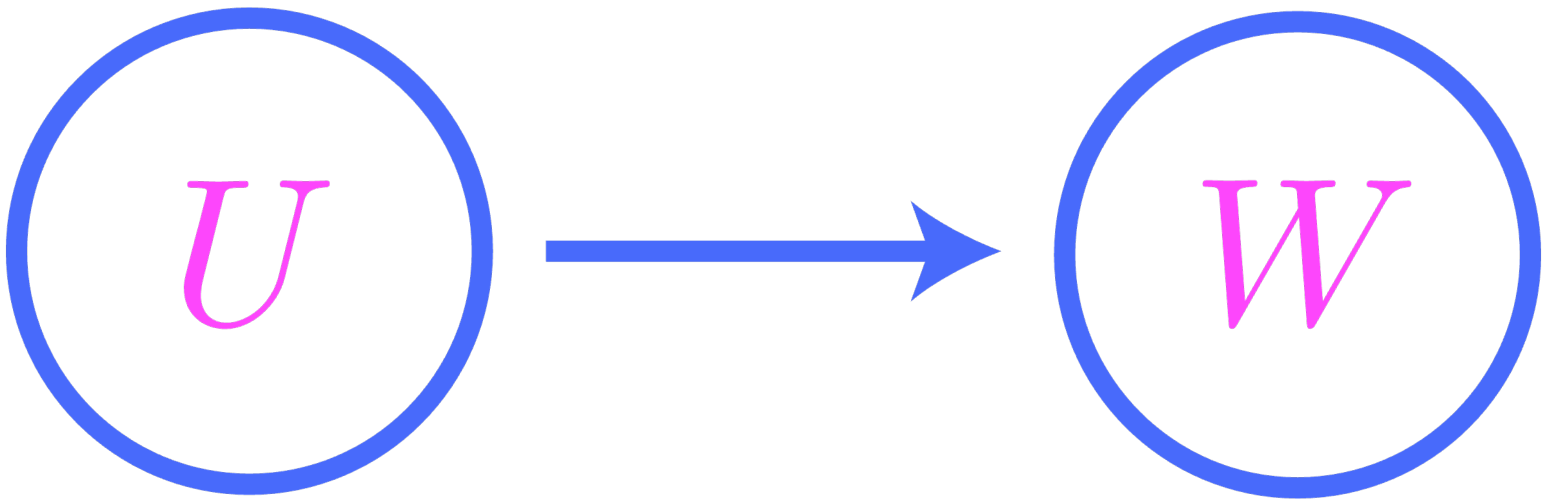
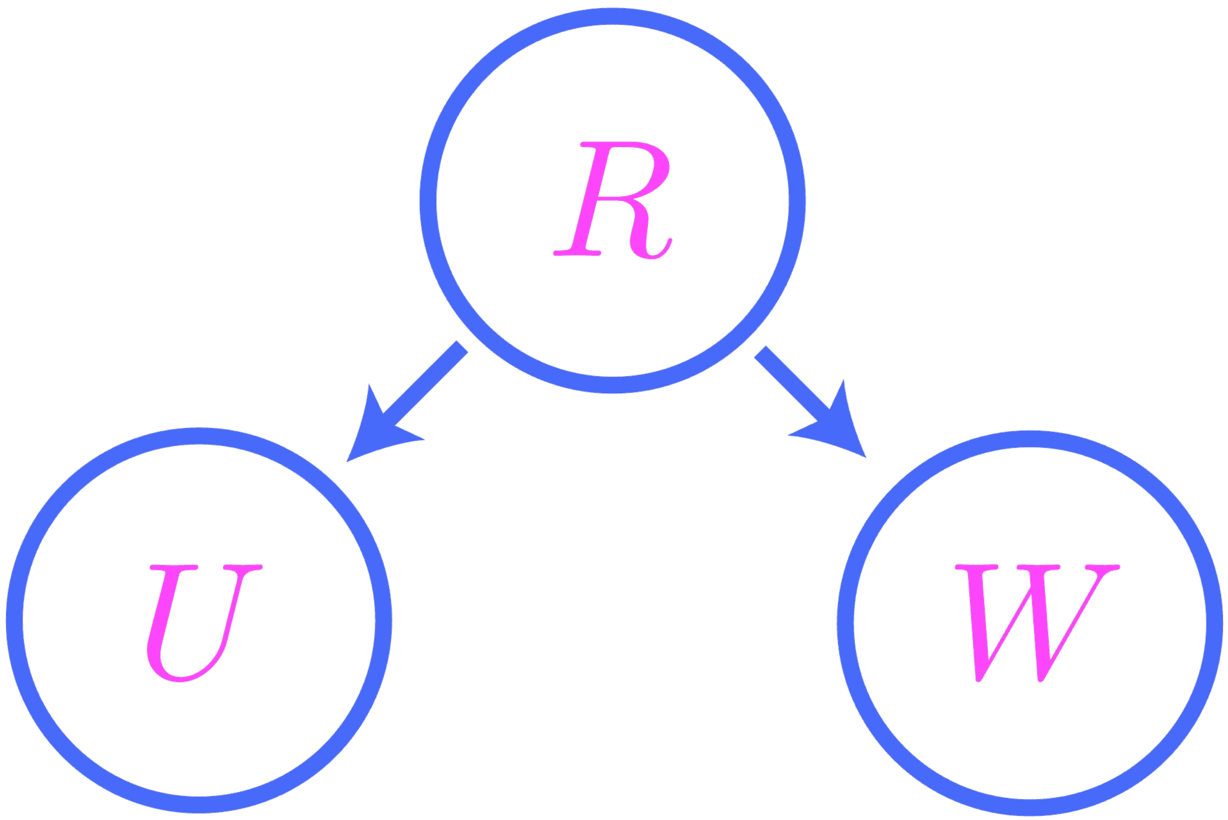
12. Probabilistic graphical models

12.1. A brief look at causal inference


X
Y


U
W


R
U
W


U
R
W

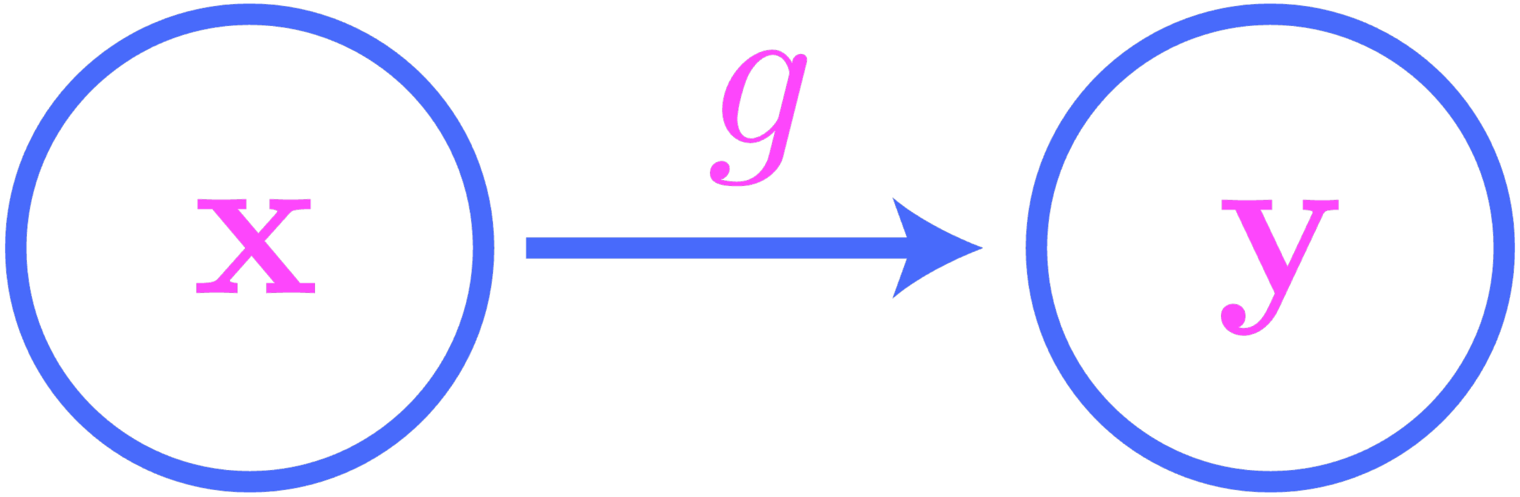
Causal structures and probability.

- 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.

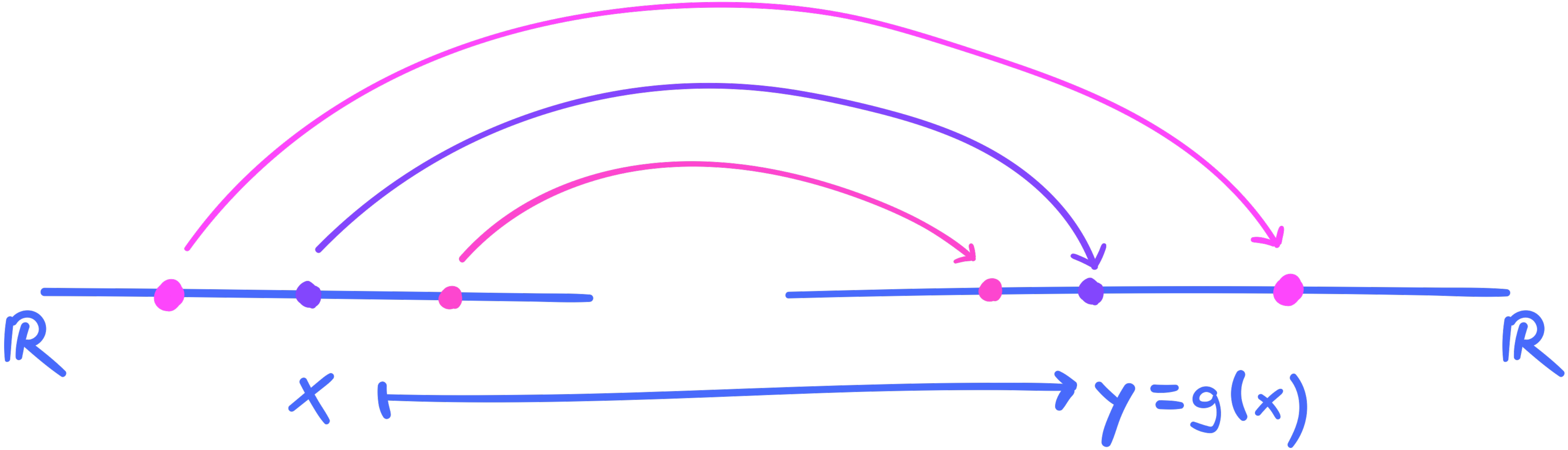
Problem Prompt

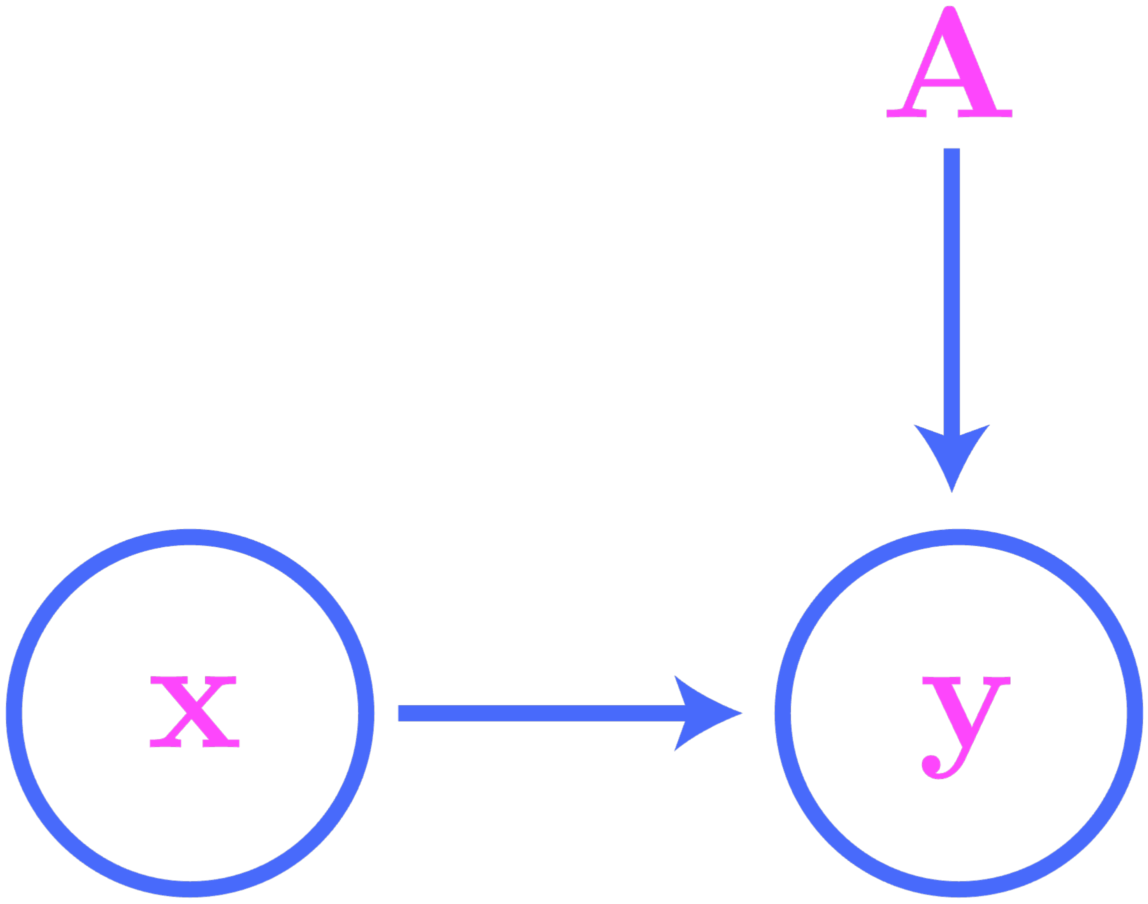
Do problem 1 on the worksheet.

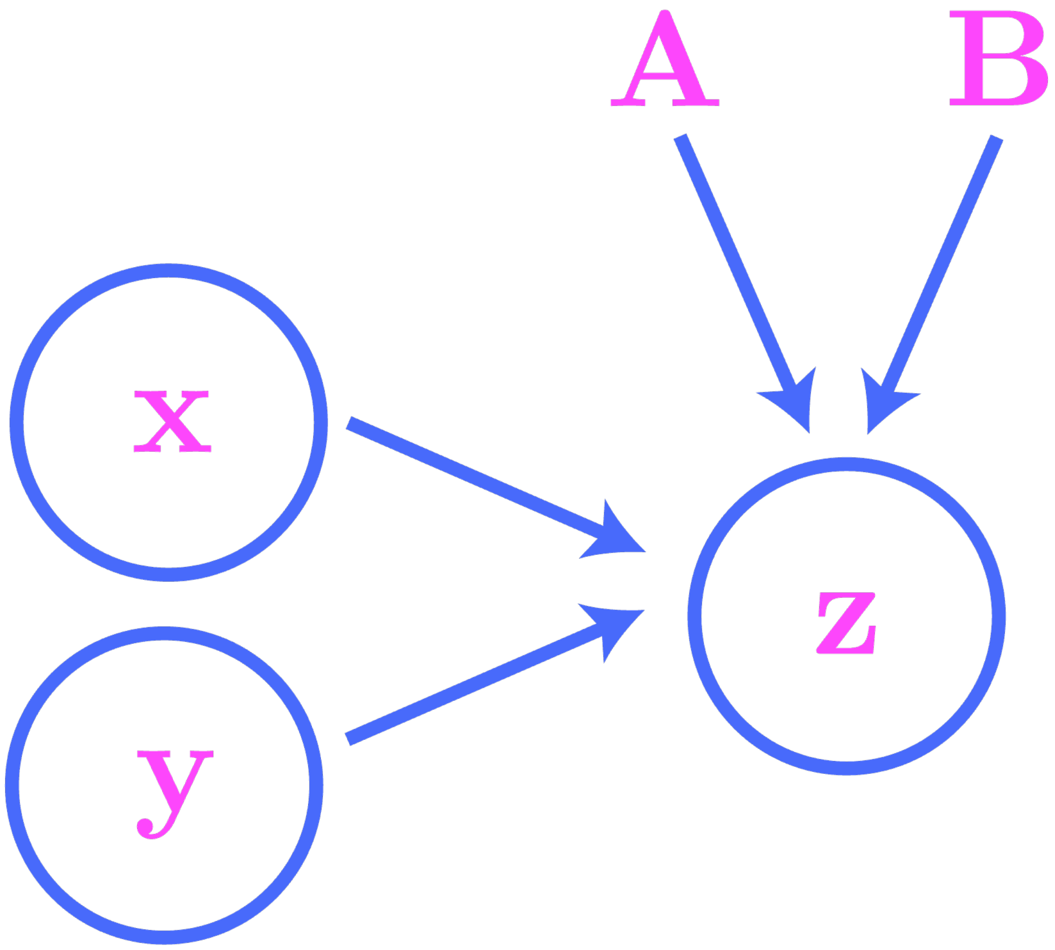
12.2. General probabilistic graphical models

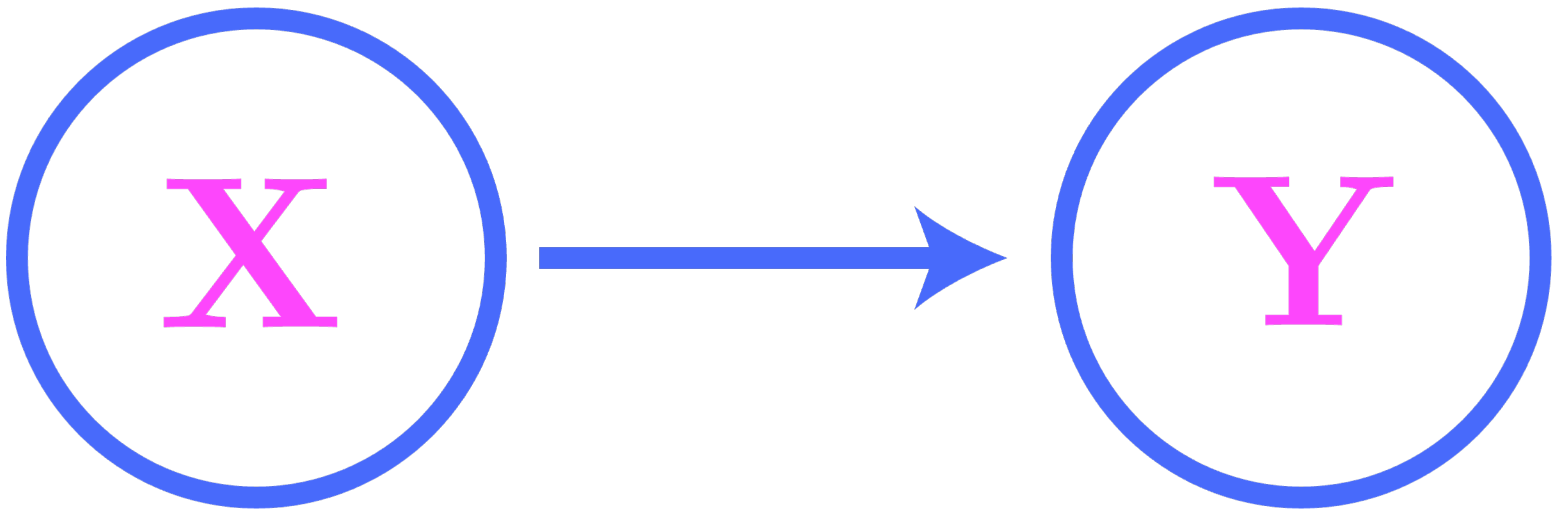

x
g
y

Deterministic link

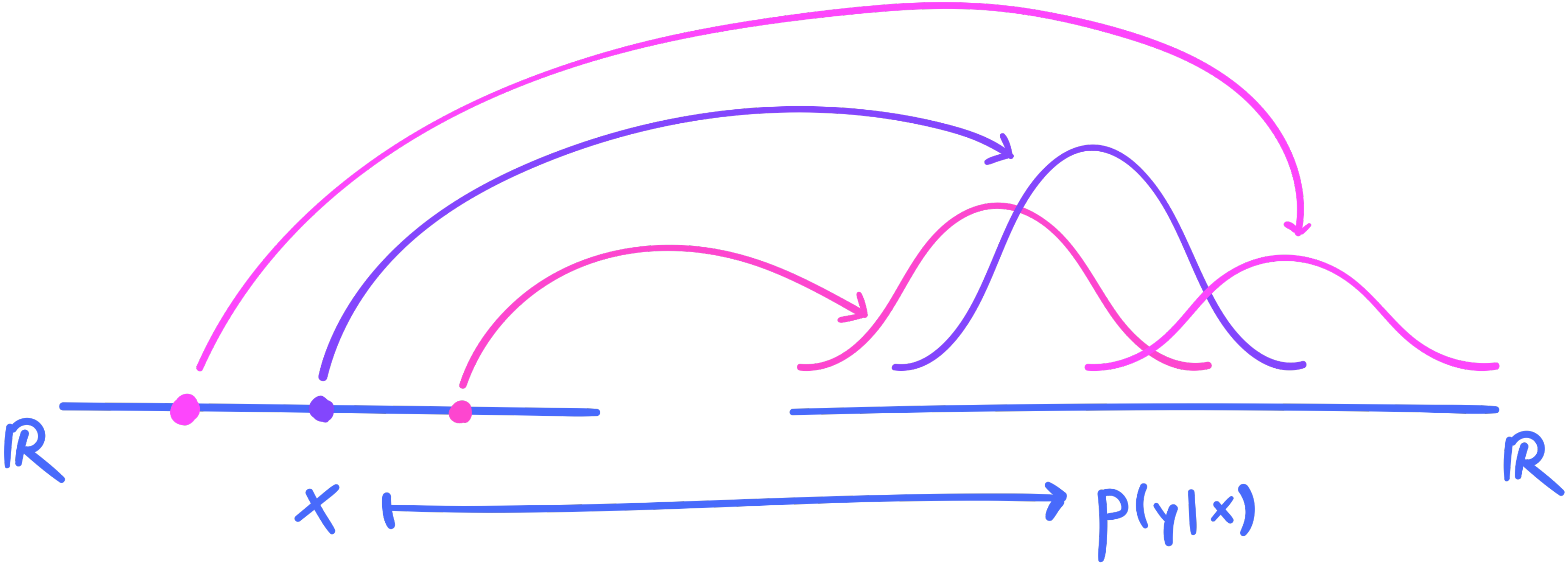



A
x
y


A
B
x
y
z

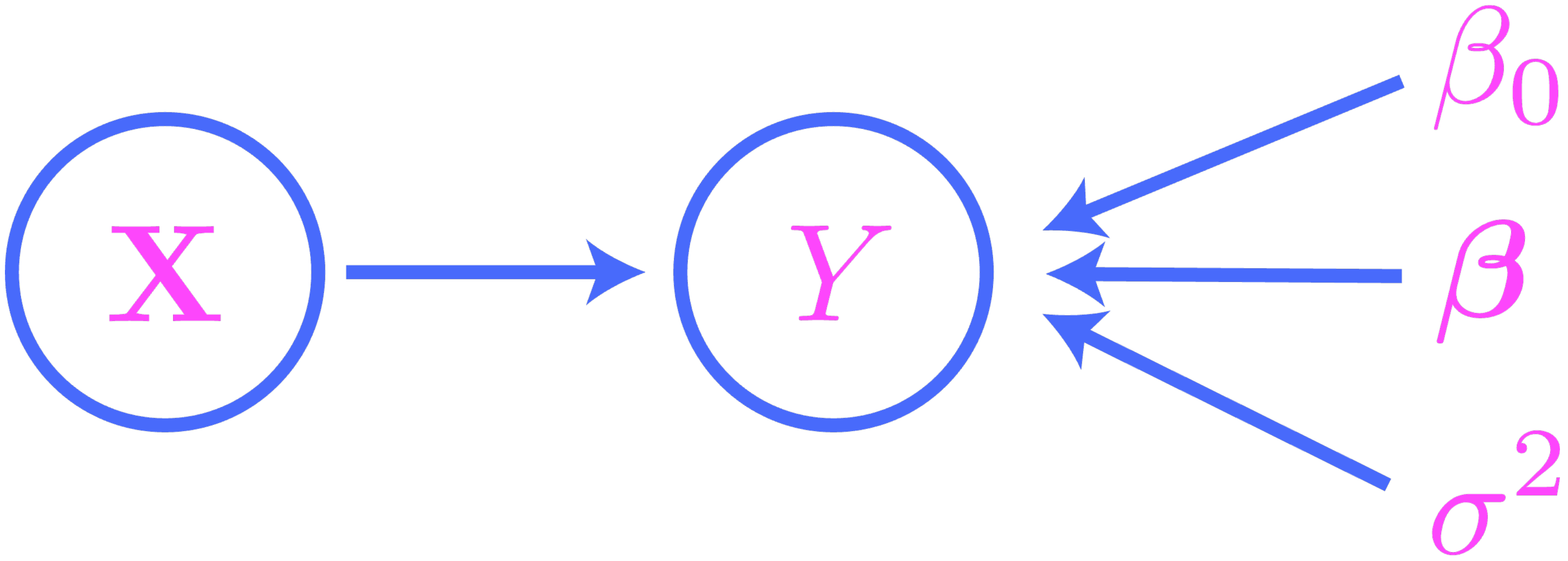

X
Y

Stochastic link (Markov Kernel)

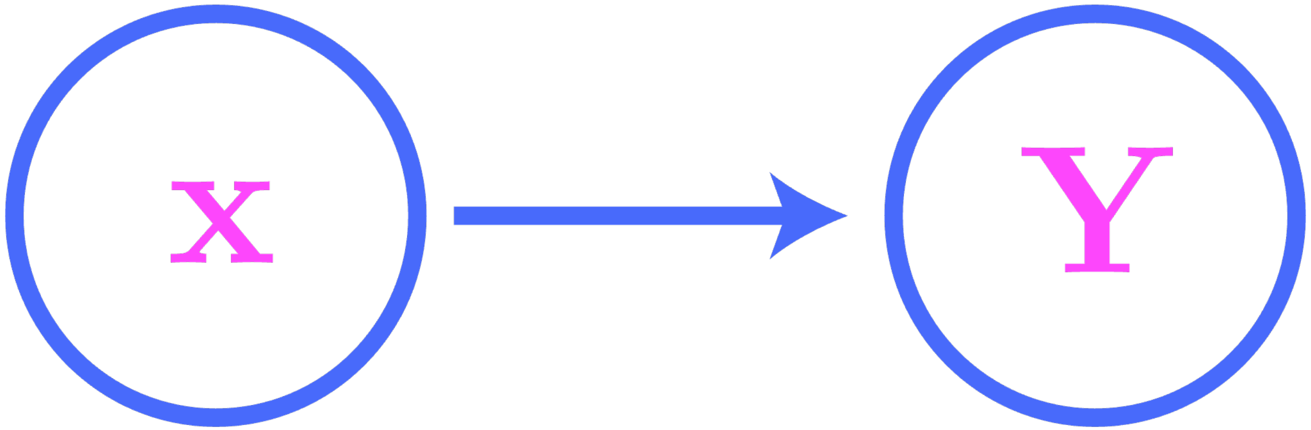


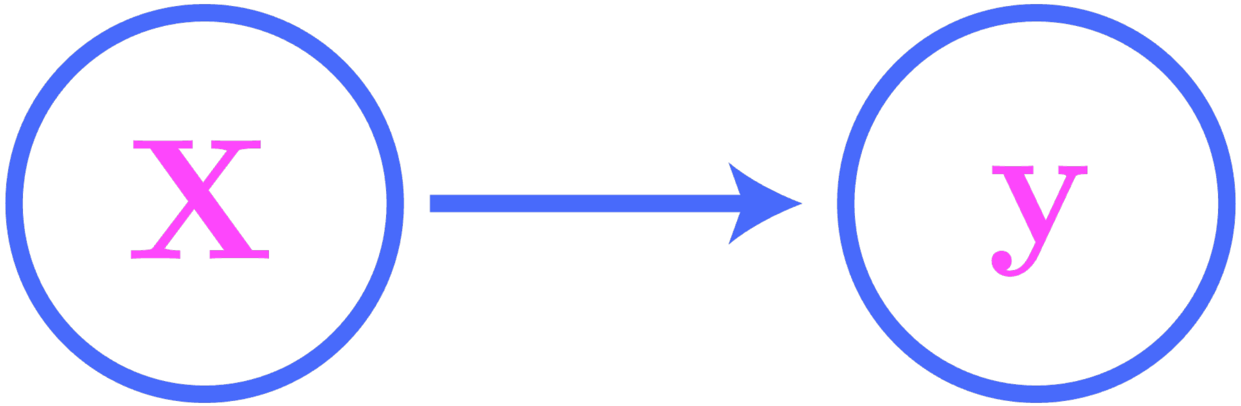
Problem Prompt

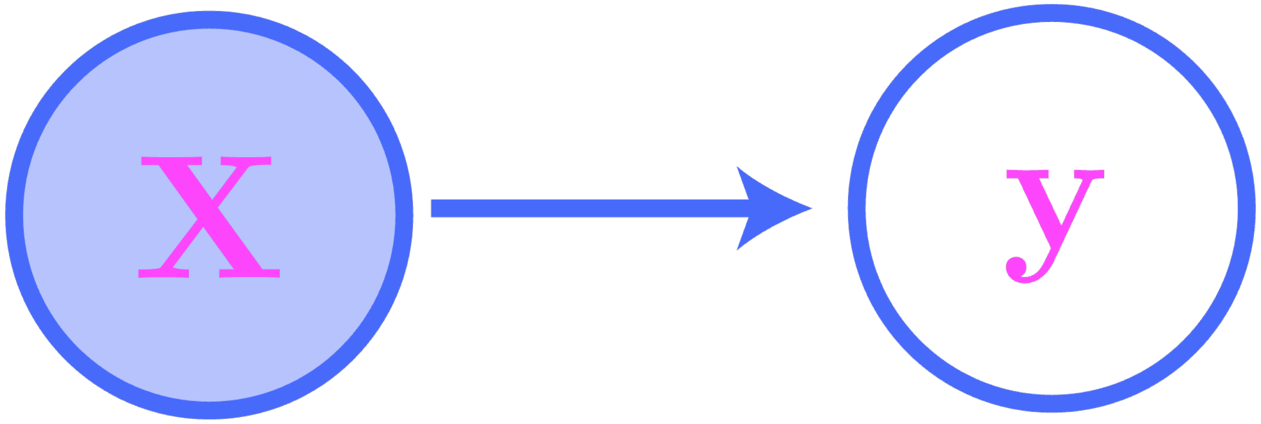
Do problem 2 on the worksheet.

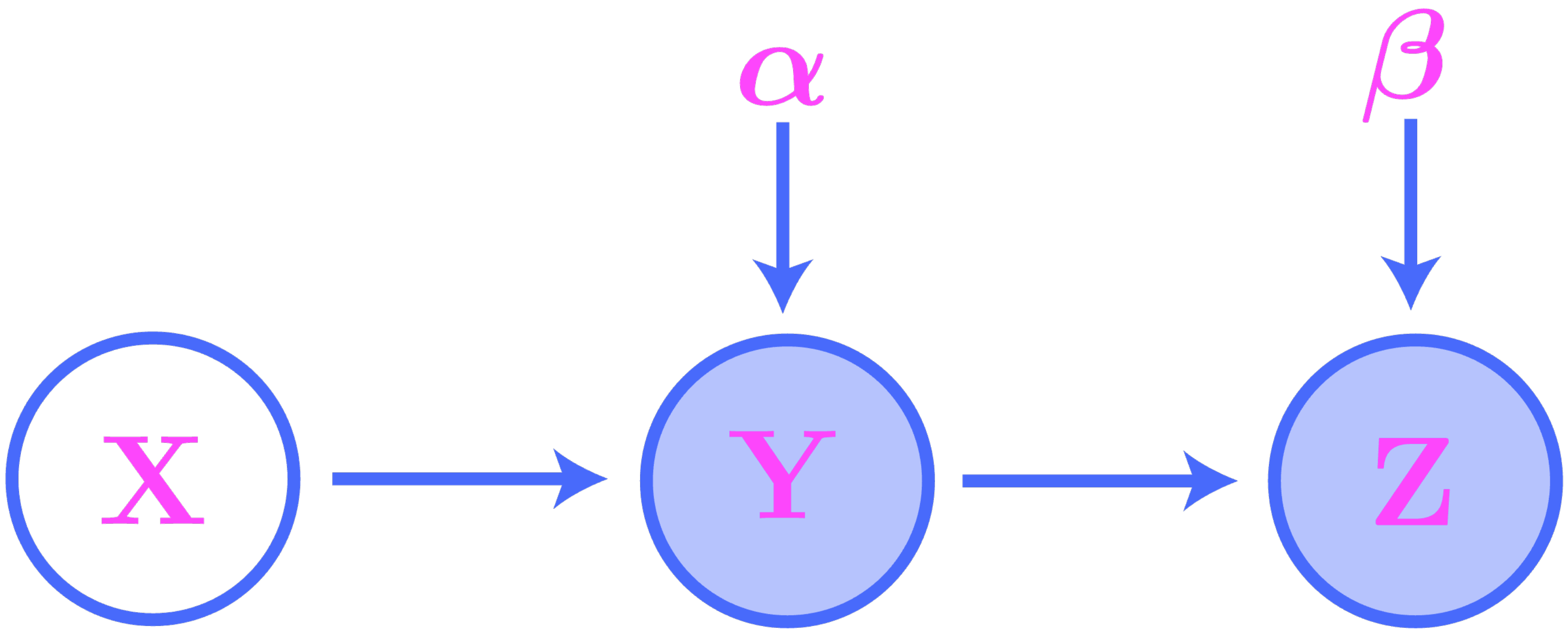


$\mathbf{X}$
Y
β_0
β
σ^2

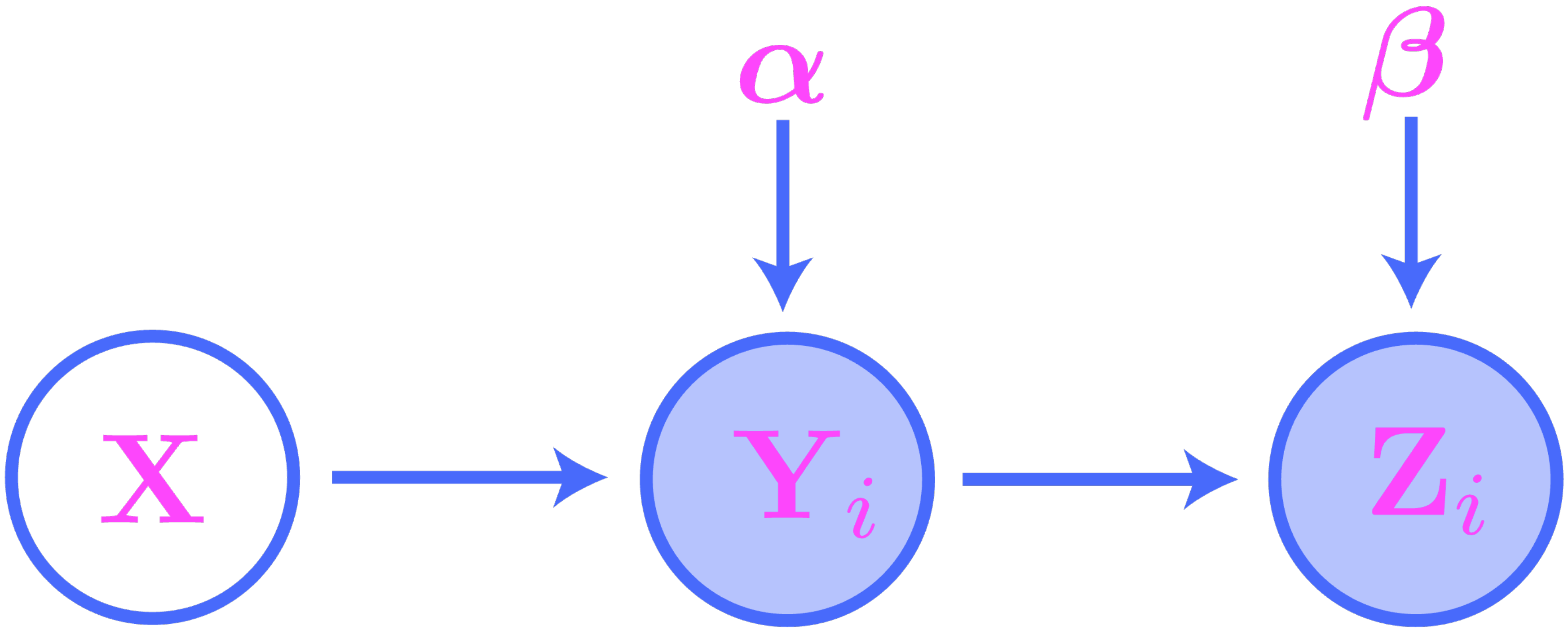

x
Y

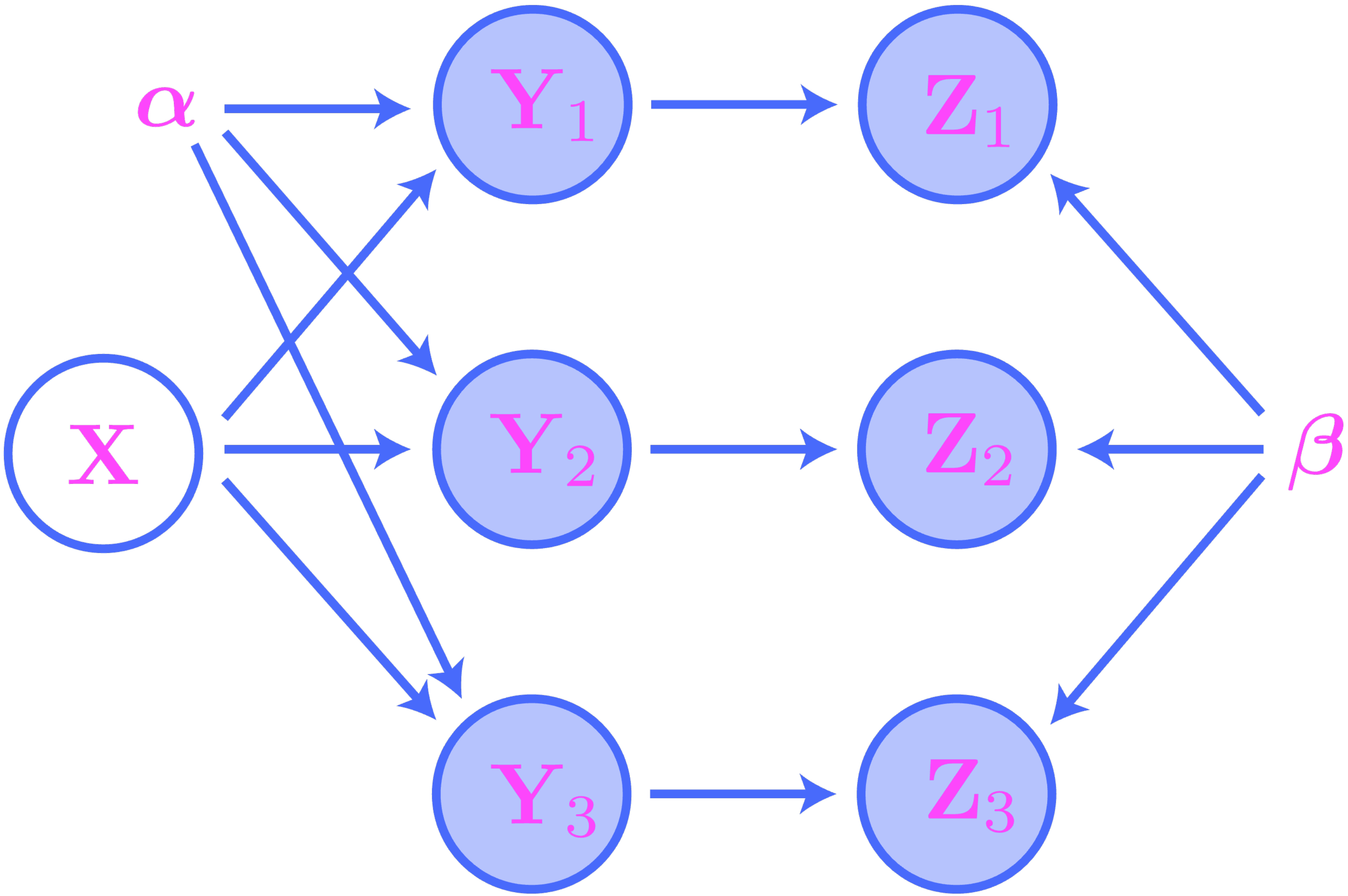

x
y


x
y

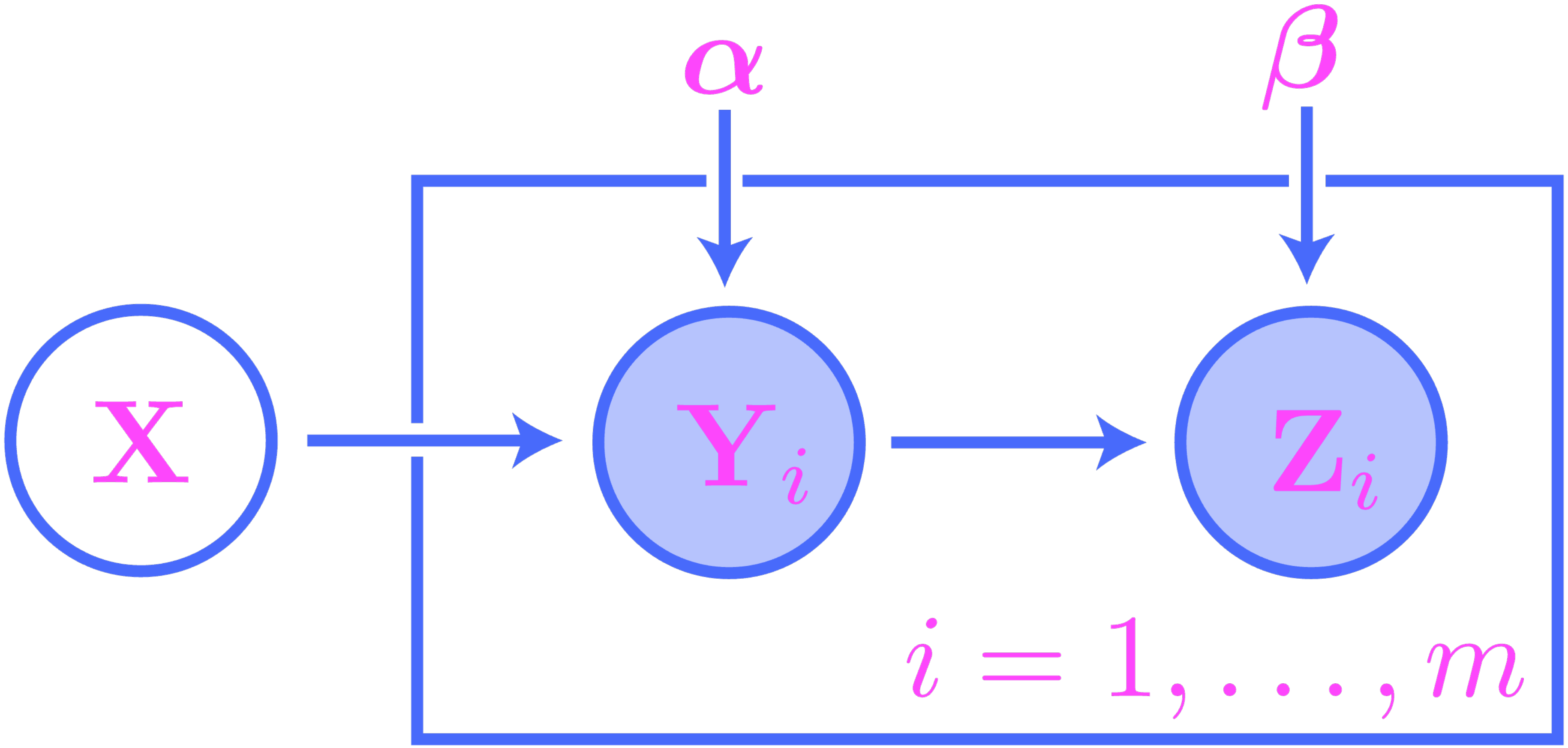


α
β
X
Y
Z


α
β
X
Y_i
Z_i



α
$\mathbf{Y}_1$
$\mathbf{Z}_1$
$\mathbf{X}$
$\mathbf{Y}_2$
$\mathbf{Z}_2$
β
$\mathbf{Y}_3$
$\mathbf{Z}_3$


$\boldsymbol{\alpha}$
$\boldsymbol{\beta}$
$\mathbf{X}$
$\mathbf{Y}_i$
$\mathbf{Z}_i$
$i = 1, \ldots, m$

Problem Prompt

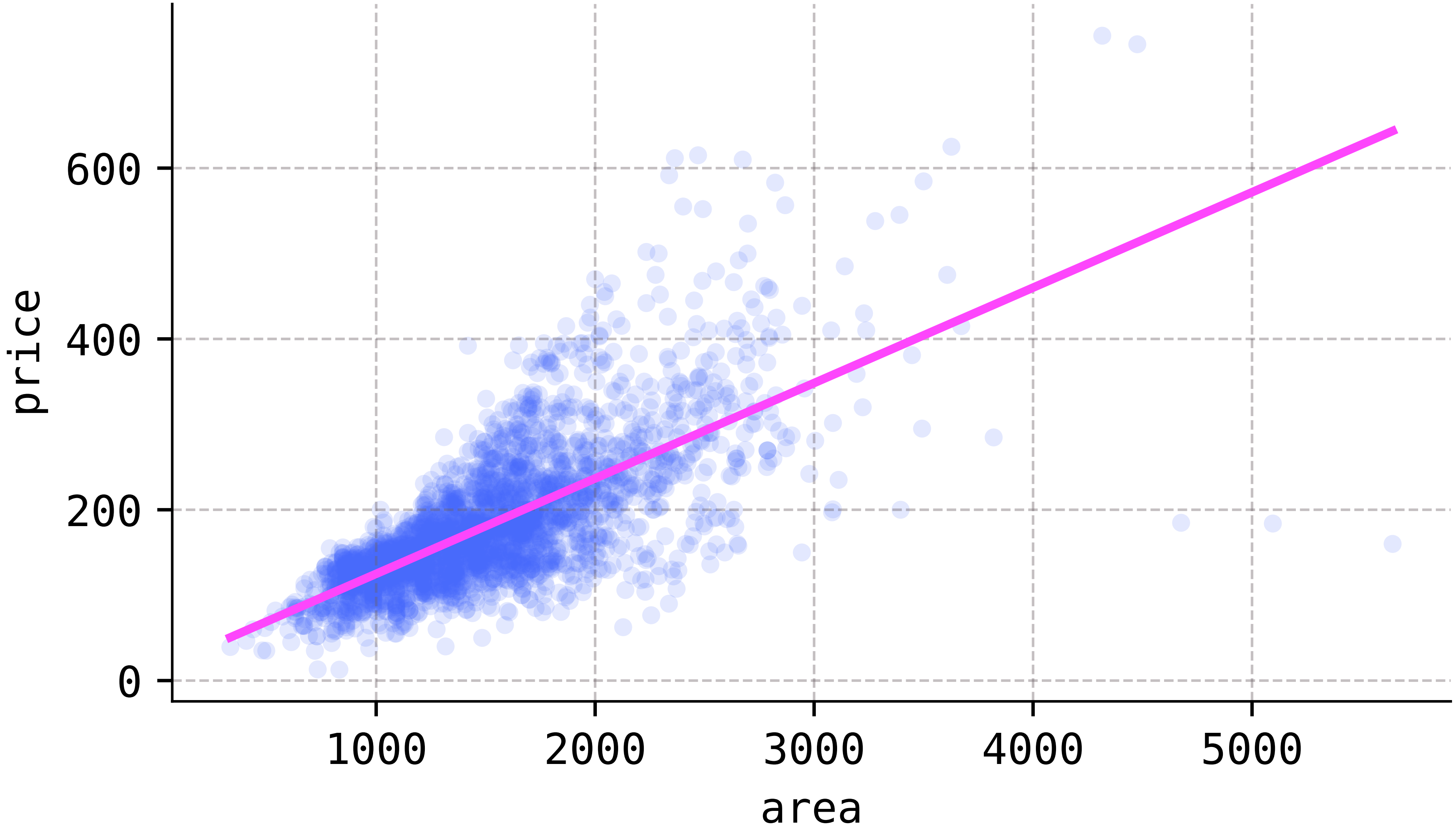
Do problem 3 on the worksheet.

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.

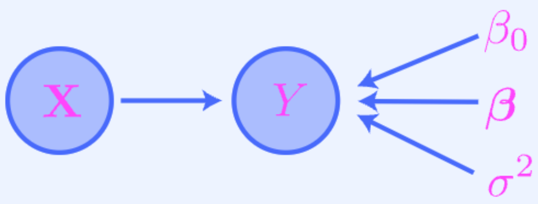
12.3. Linear regression models



price
600
400
200
0
1000
2000
3000
4000
5000
area

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 $\boldsymbol{\beta} \in \mathbb{R}^n$.
- A positive real parameter $\sigma^2 > 0$.

The link function at Y is given by

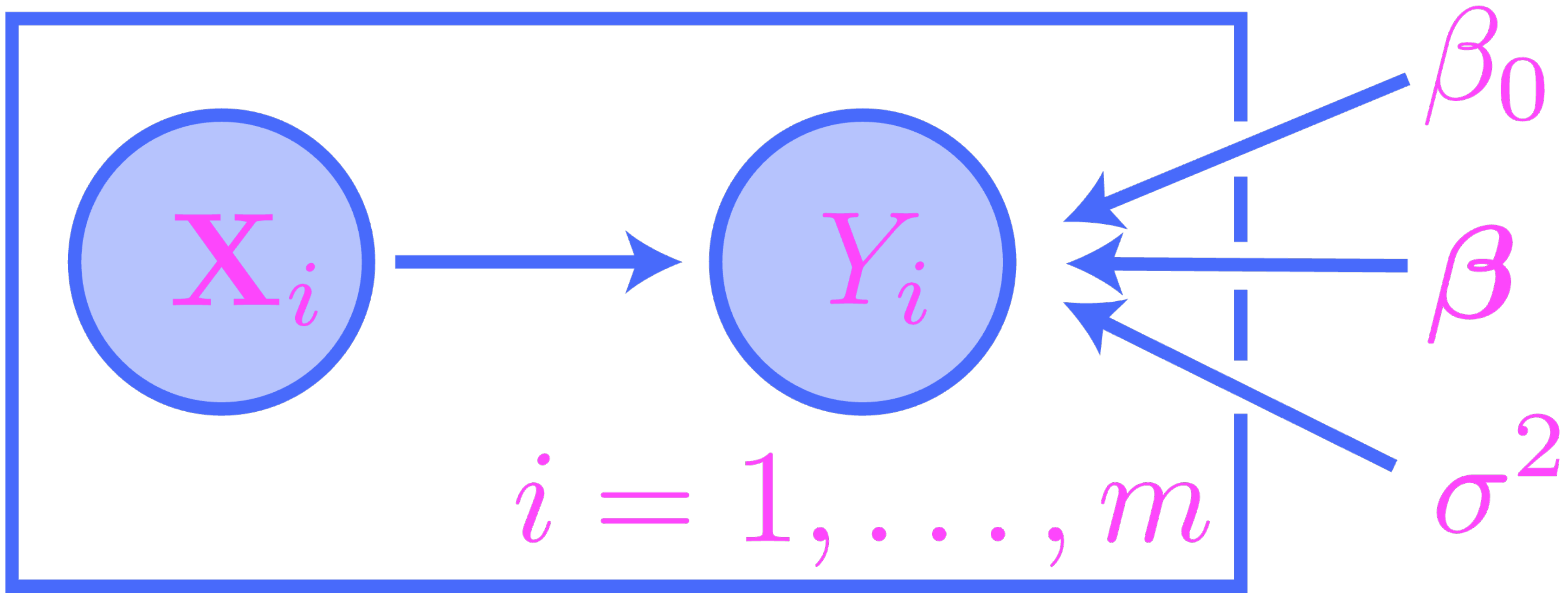
$$Y \mid \mathbf{X} \sim \mathcal{N}(\mu, \sigma^2), \quad \text{where} \quad \mu = \beta_0 + \mathbf{x}^\intercal \boldsymbol{\beta}.$$

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

$$\mathcal{L}(\beta_0, \boldsymbol{\beta}, \sigma^2;\ y \mid \mathbf{x}) \stackrel{\text{def}}{=} f\big(y \mid \mathbf{x};\ \beta_0, \boldsymbol{\beta}, \sigma^2\big) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right] \tag{12.2}$$

of the parameters $\beta_0, \boldsymbol{\beta}, \sigma^2$, where $\mu = \beta_0 + \mathbf{x}^\intercal \boldsymbol{\beta}$.


$\mathbf{X}_i$
Y_i
$i = 1, \ldots, m$
β_0
$\boldsymbol{\beta}$
σ^2

Definition 12.4

Given an observed dataset

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

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

$$\mathcal{L}(\beta_0, \boldsymbol{\beta}, \sigma^2; y_1, \ldots, y_m \mid \mathbf{x}_1, \ldots, \mathbf{x}_m) \stackrel{\text{def}}{=} f\big(y_1, \ldots, y_m \mid \mathbf{x}_1, \ldots, \mathbf{x}_m;\ \beta_0, \boldsymbol{\beta}, \sigma^2\big)$$

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

Theorem 12.1 (Data likelihood functions of linear regression models)

Given an observed dataset

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

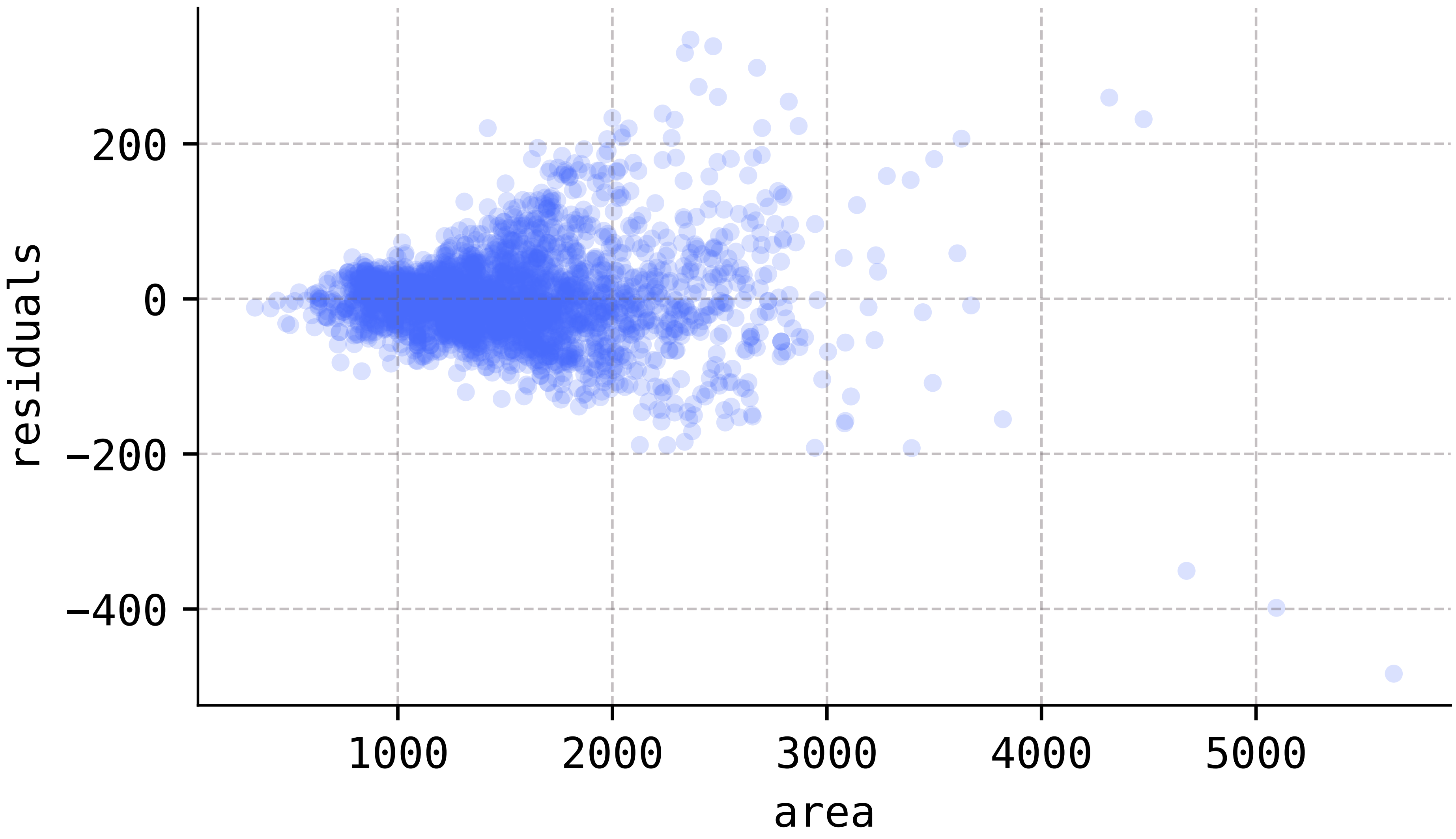
the data likelihood function for a linear regression model is given by

$$\begin{aligned}
\mathcal{L}_{\text{data}}(\beta_0, \boldsymbol{\beta}, \sigma^2) &= \prod_{i=1}^{m} \mathcal{L}(\beta_0, \boldsymbol{\beta}, \sigma^2;\ y_i \mid \mathbf{x}_i) \\
&= \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i - \mu_i)^2\right],
\end{aligned}$$

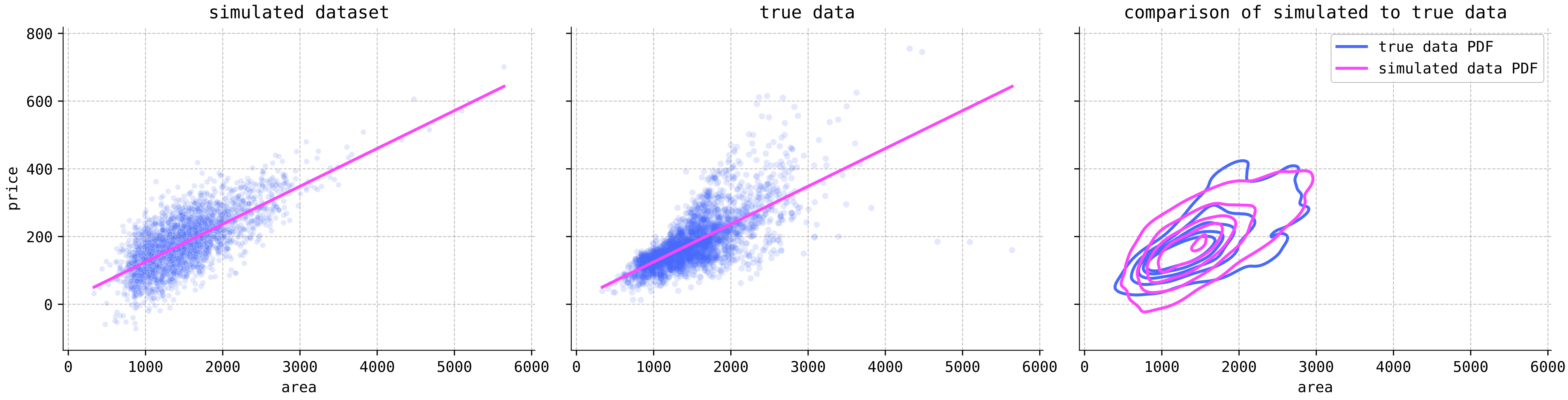
where $\mu_i = \beta_0 + \mathbf{x}_i^\intercal \boldsymbol{\beta}$ for each $i = 1, \ldots, m$.

Problem Prompt

Do problem 3 on the worksheet.



200
0
−200
−400
residuals
1000
2000
3000
4000
5000
area


simulated dataset
true data
comparison of simulated to true data
price
area
true data PDF
simulated data PDF