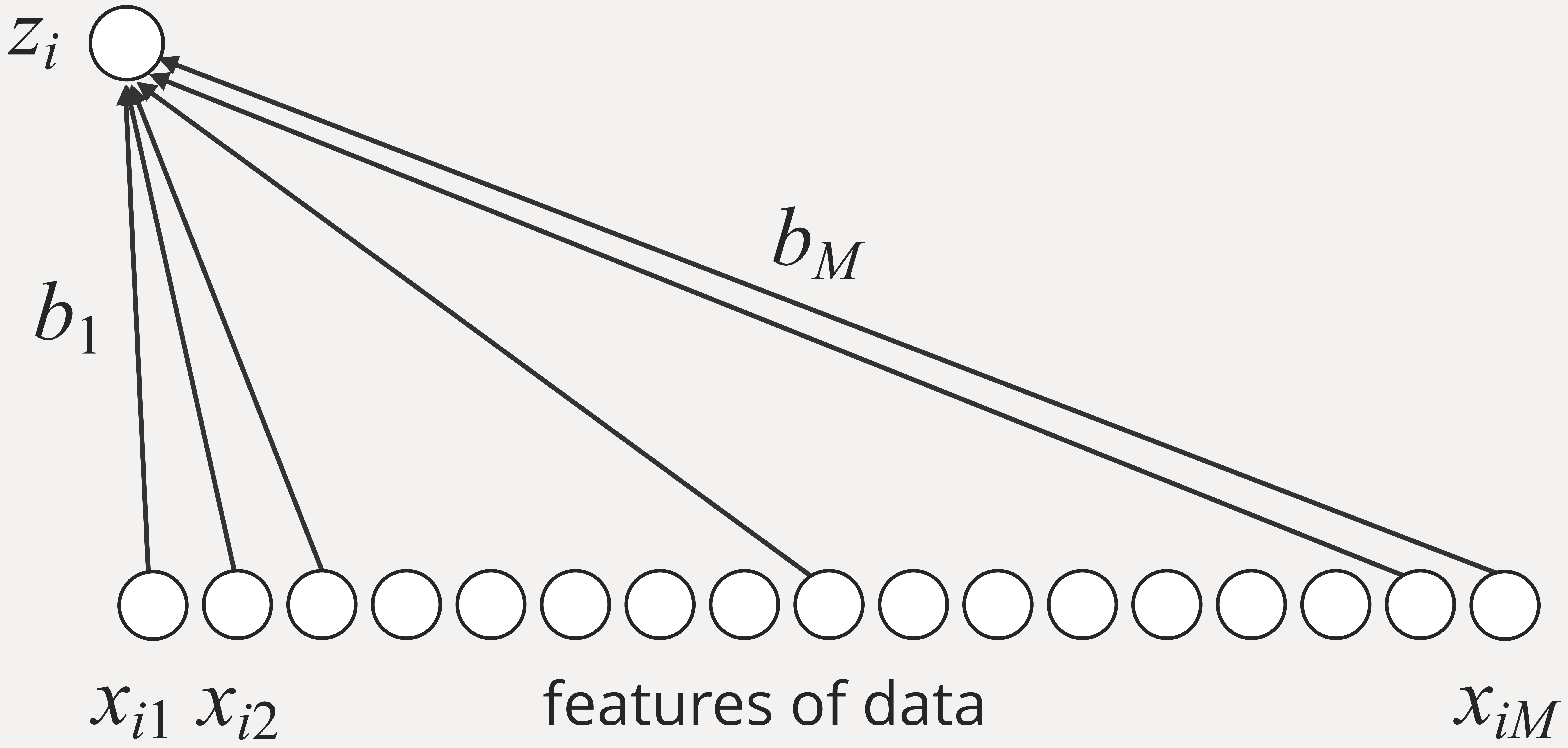
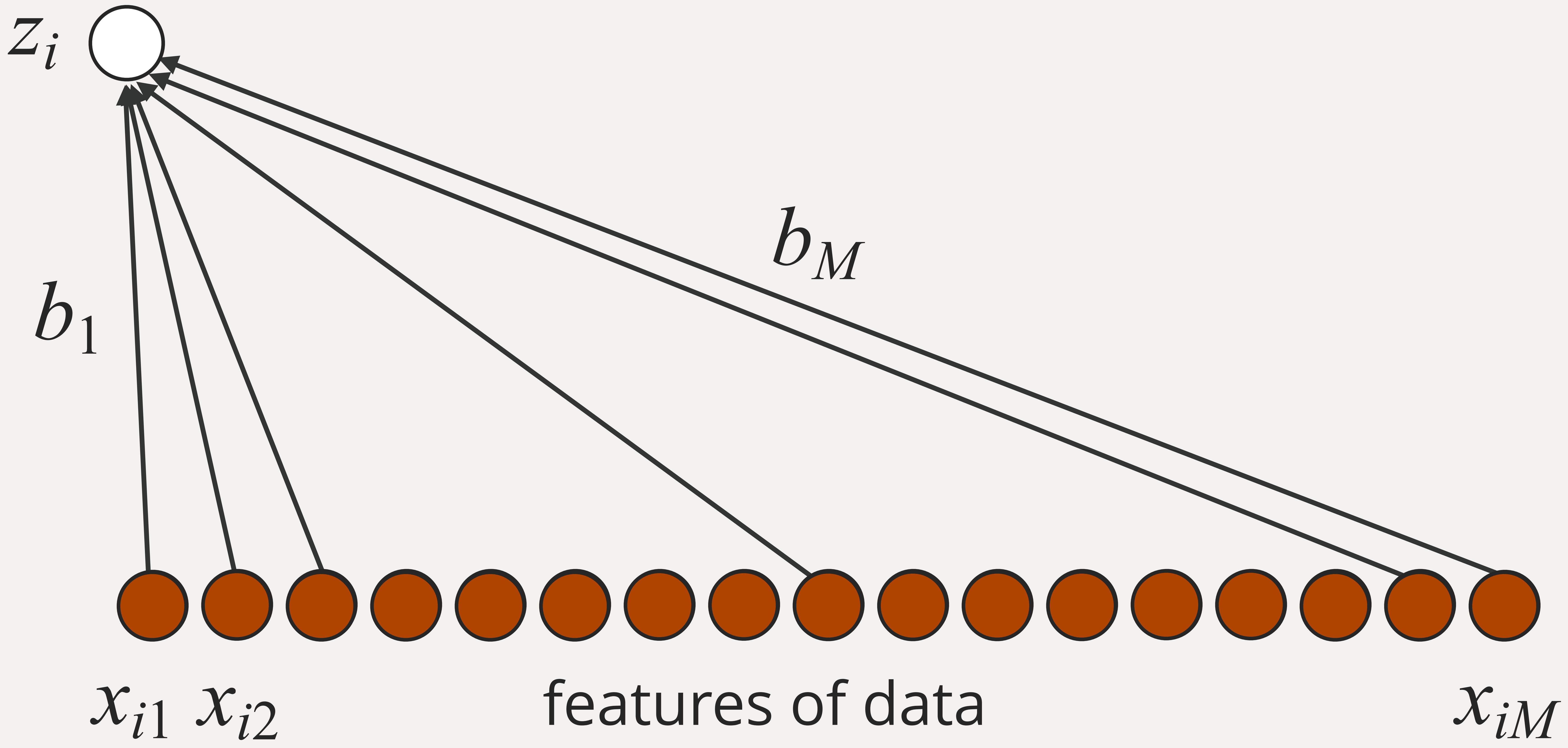
Multilayer Perceptron Concepts

Logistic Regression

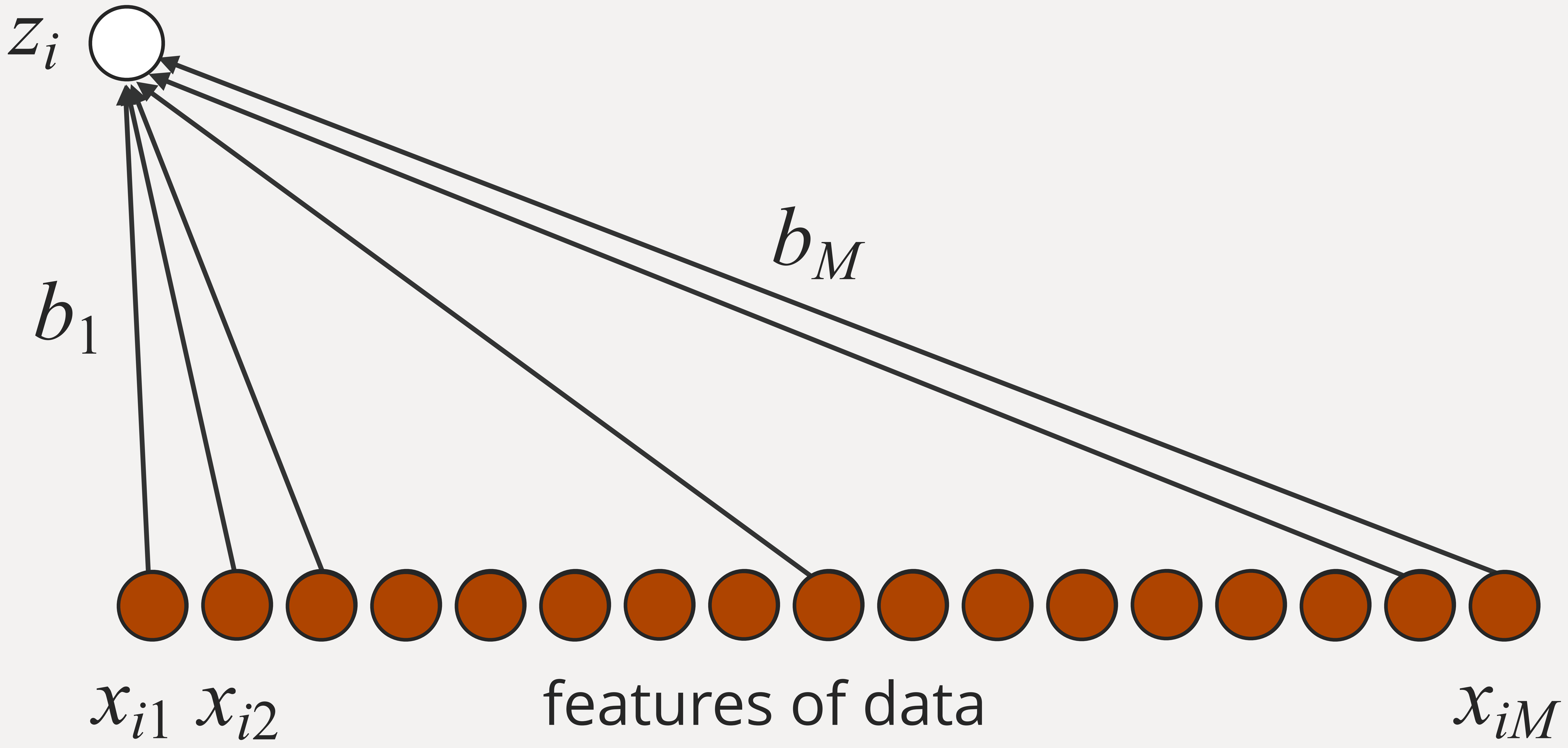


Logistic Regression



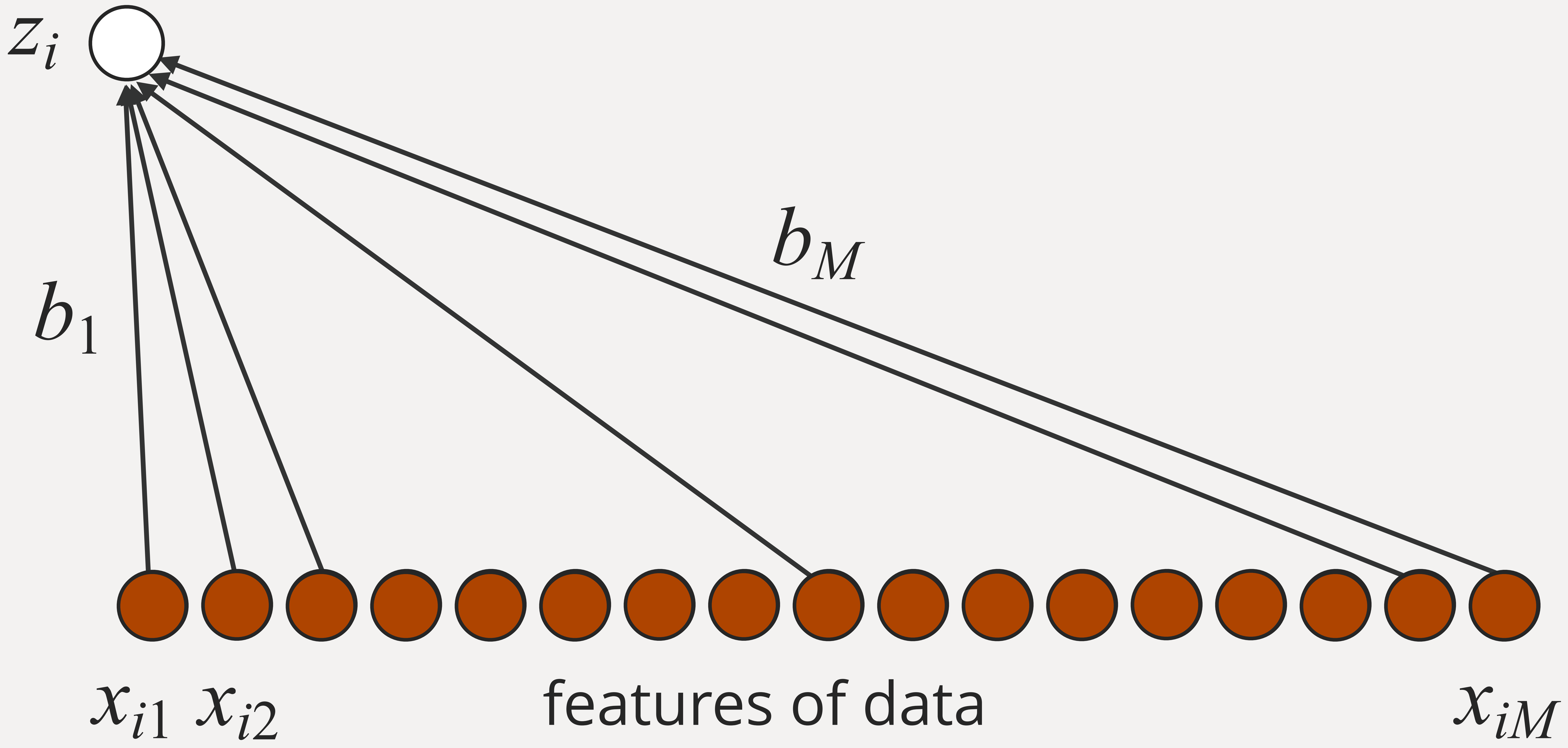
Logistic Regression

$$z_i = (b_1 \times x_{i1}) + (b_2 \times x_{i2}) + \cdots + (b_M \times x_{iM}) + b_0$$



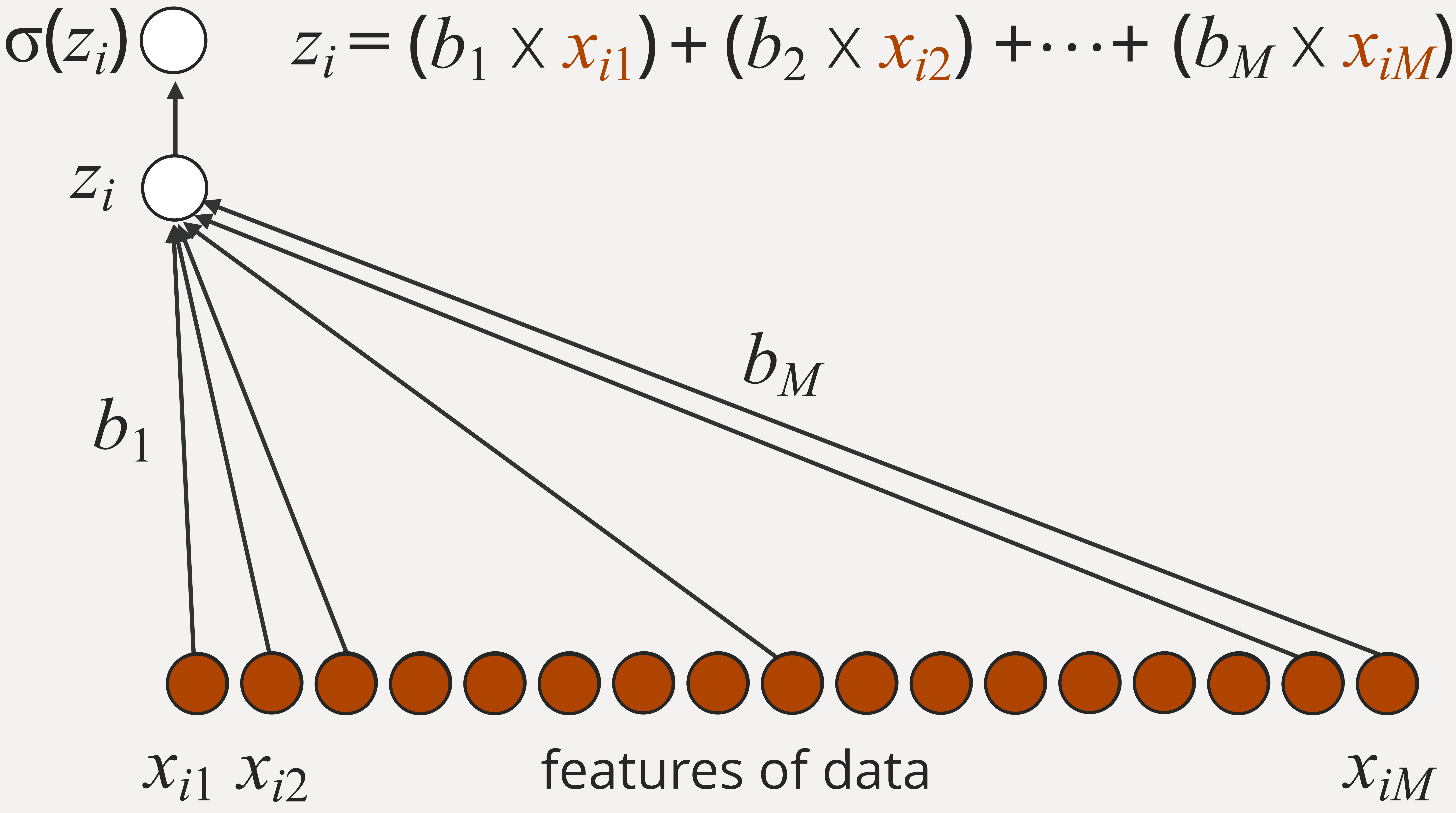
Logistic Regression

$$z_i = (b_1 \times x_{i1}) + (b_2 \times x_{i2}) + \cdots + (b_M \times x_{iM}) + b_0$$

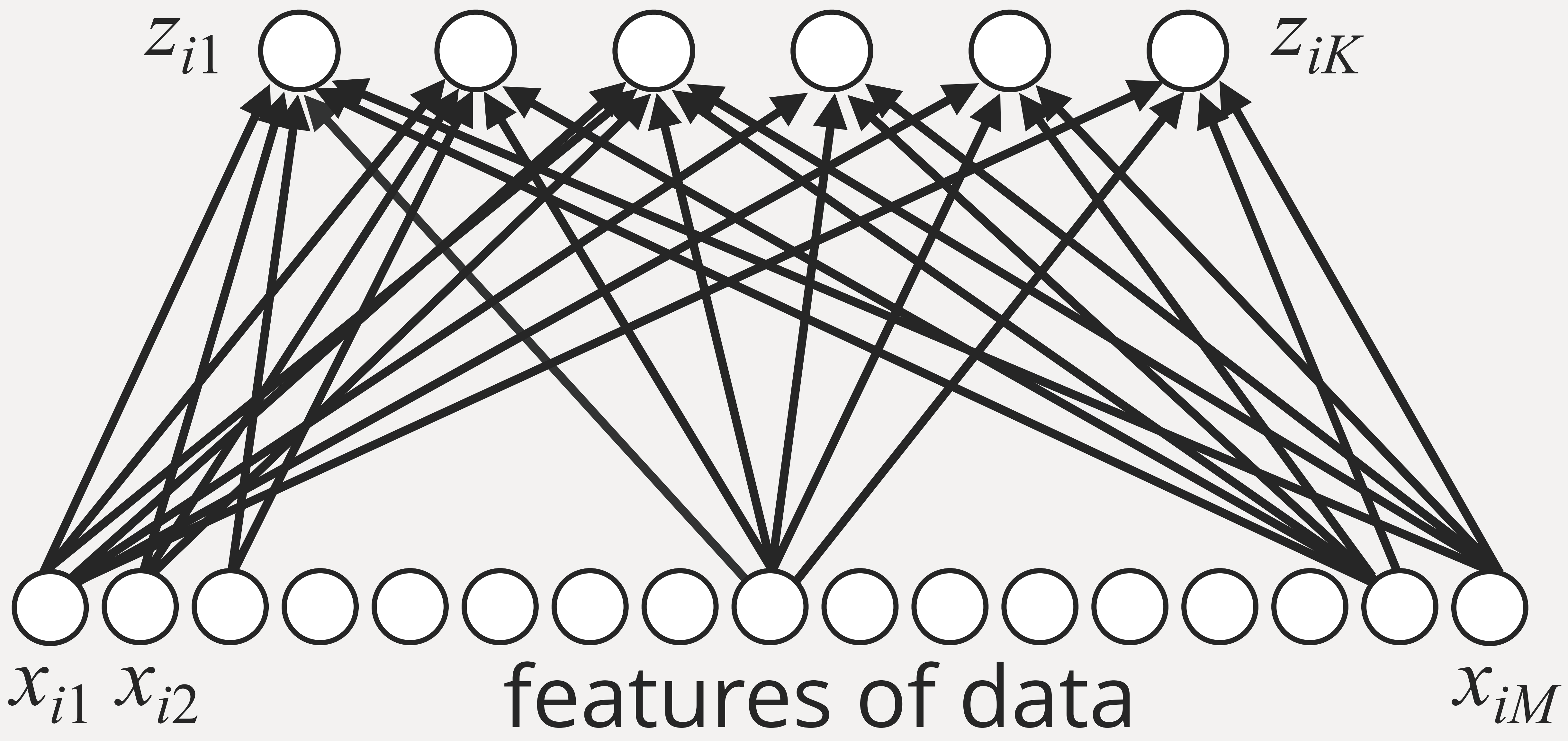


Logistic Regression

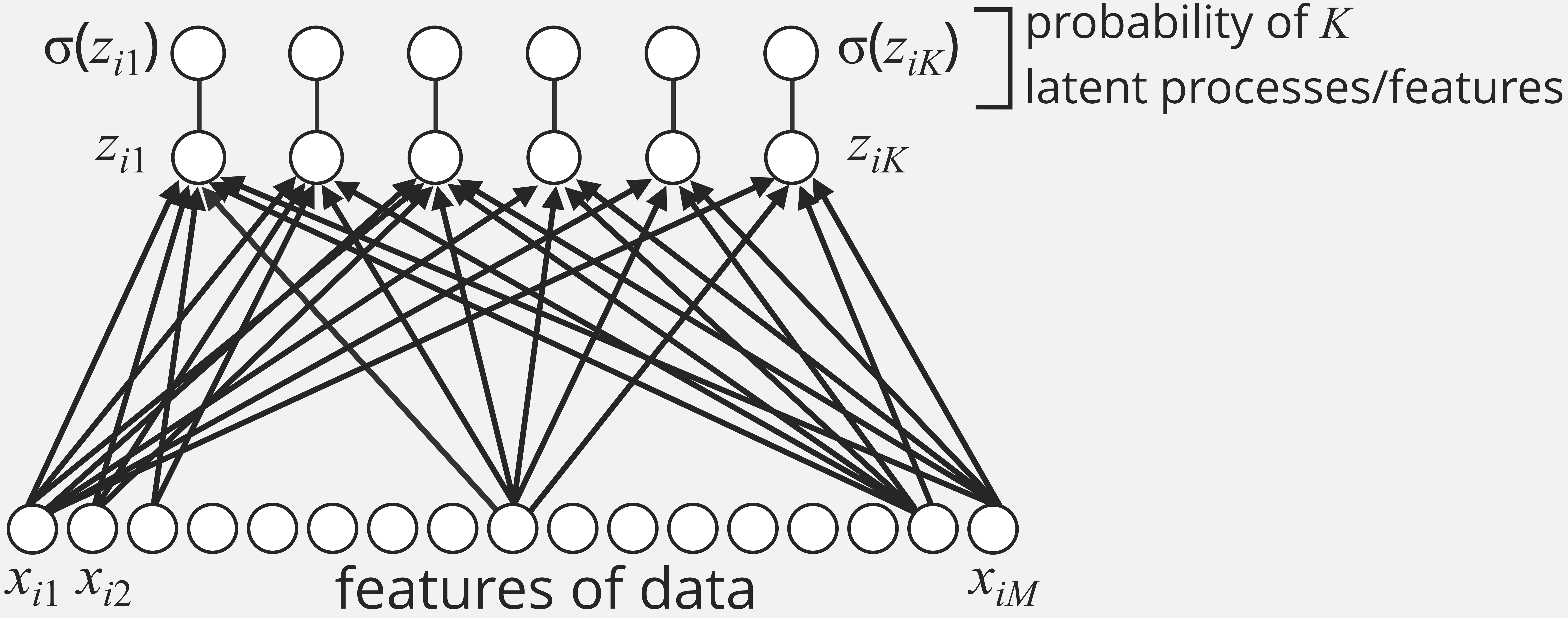
$$z_i = (b_1 \times x_{i1}) + (b_2 \times x_{i2}) + \cdots + (b_M \times x_{iM}) + b_0$$



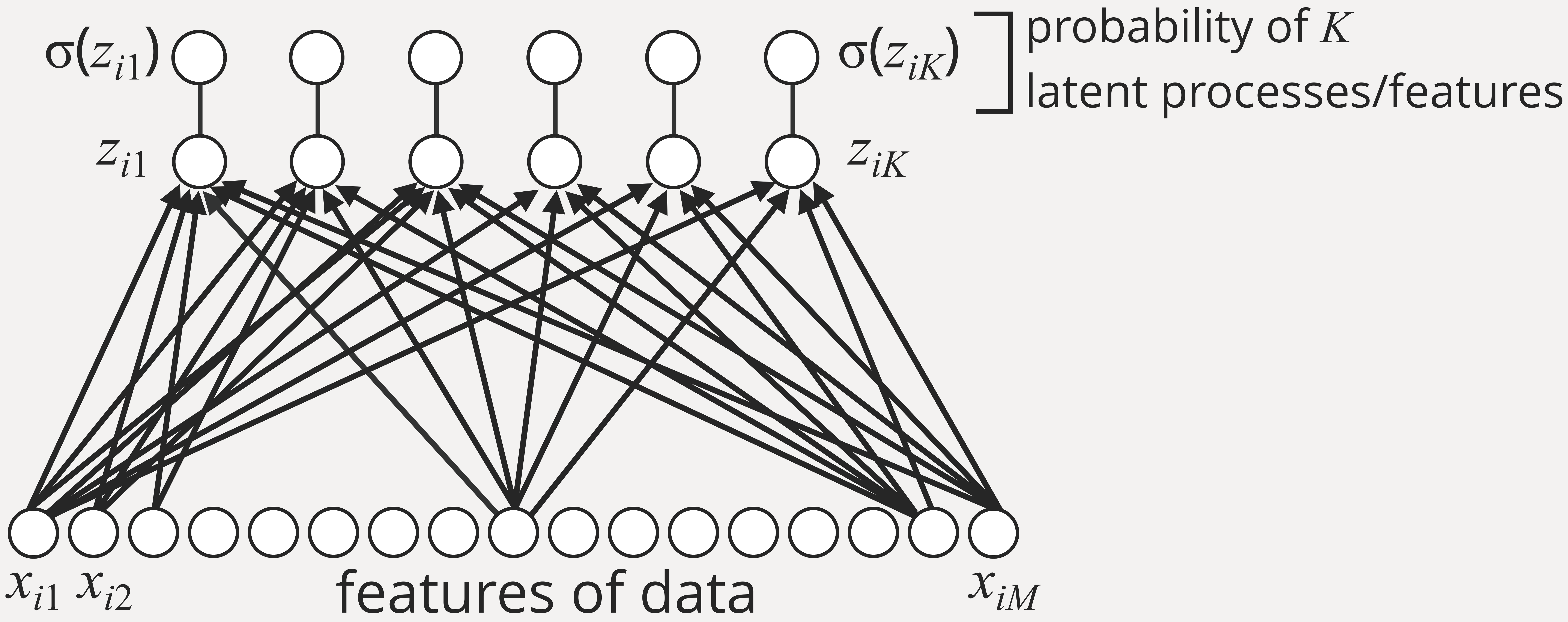
Generalization of Logistic Regression: Learned Features



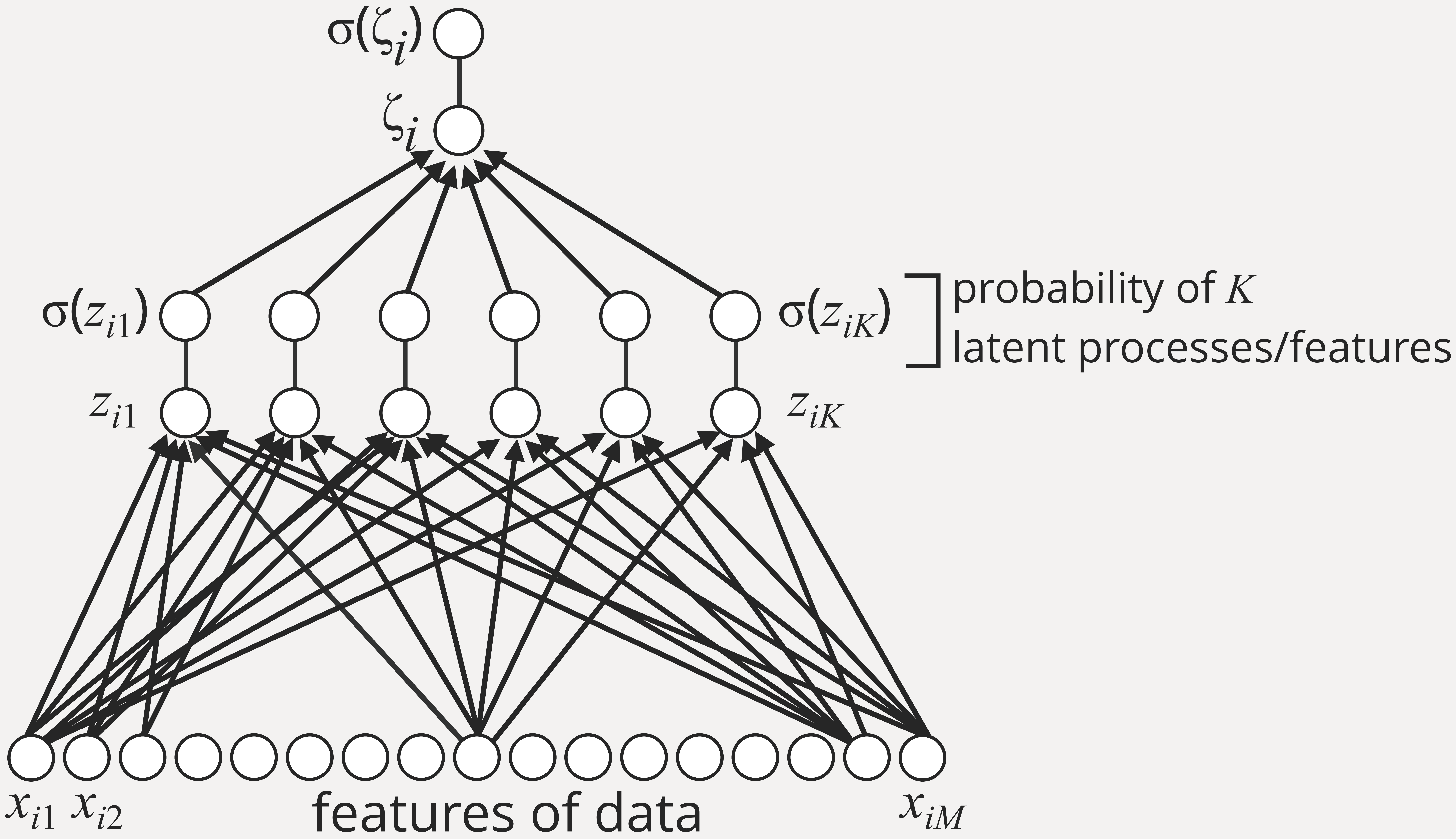
Generalization of Logistic Regression: Learned Features



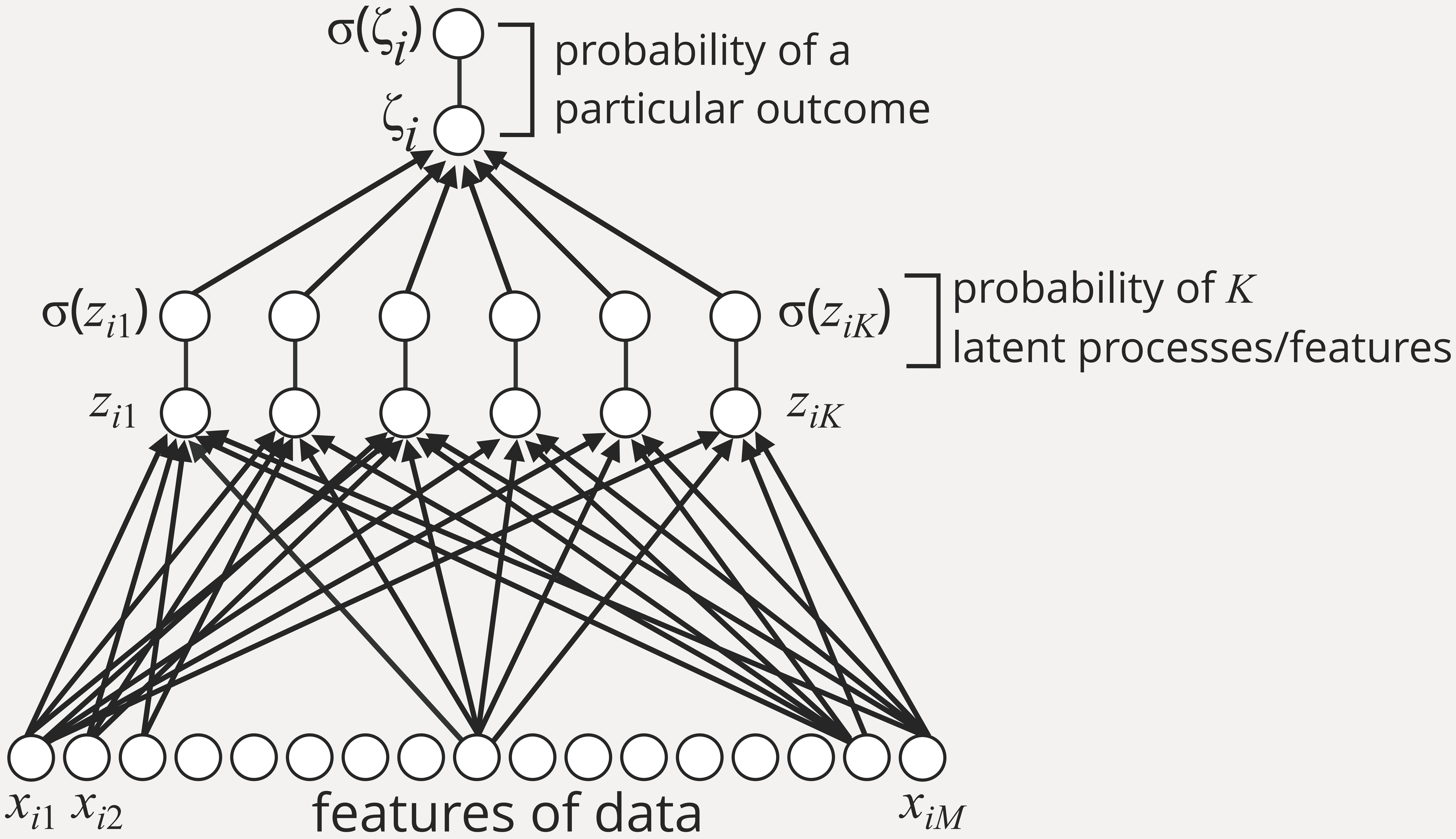
Extended Logistic Regression



Extended Logistic Regression

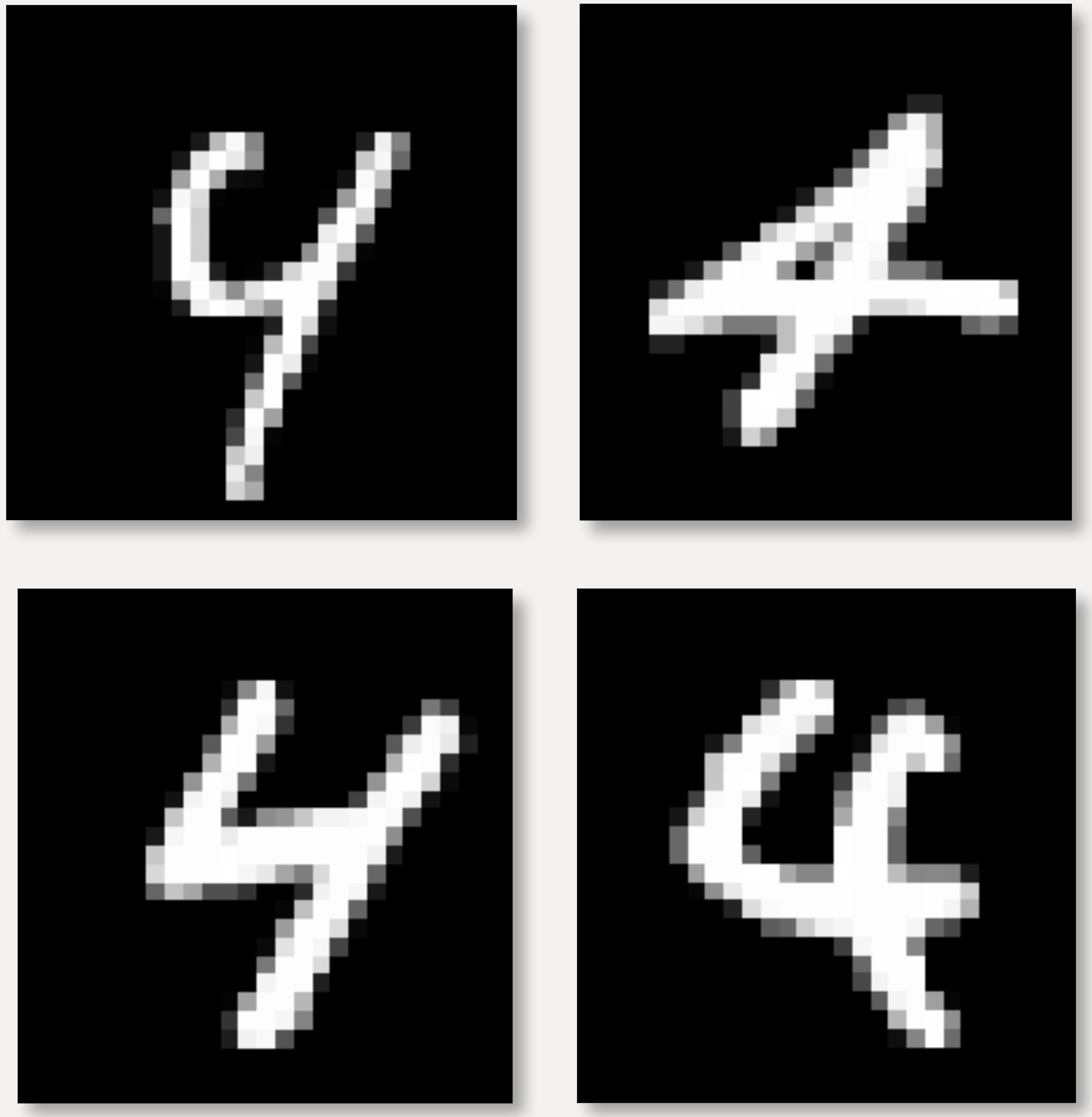


Extended Logistic Regression



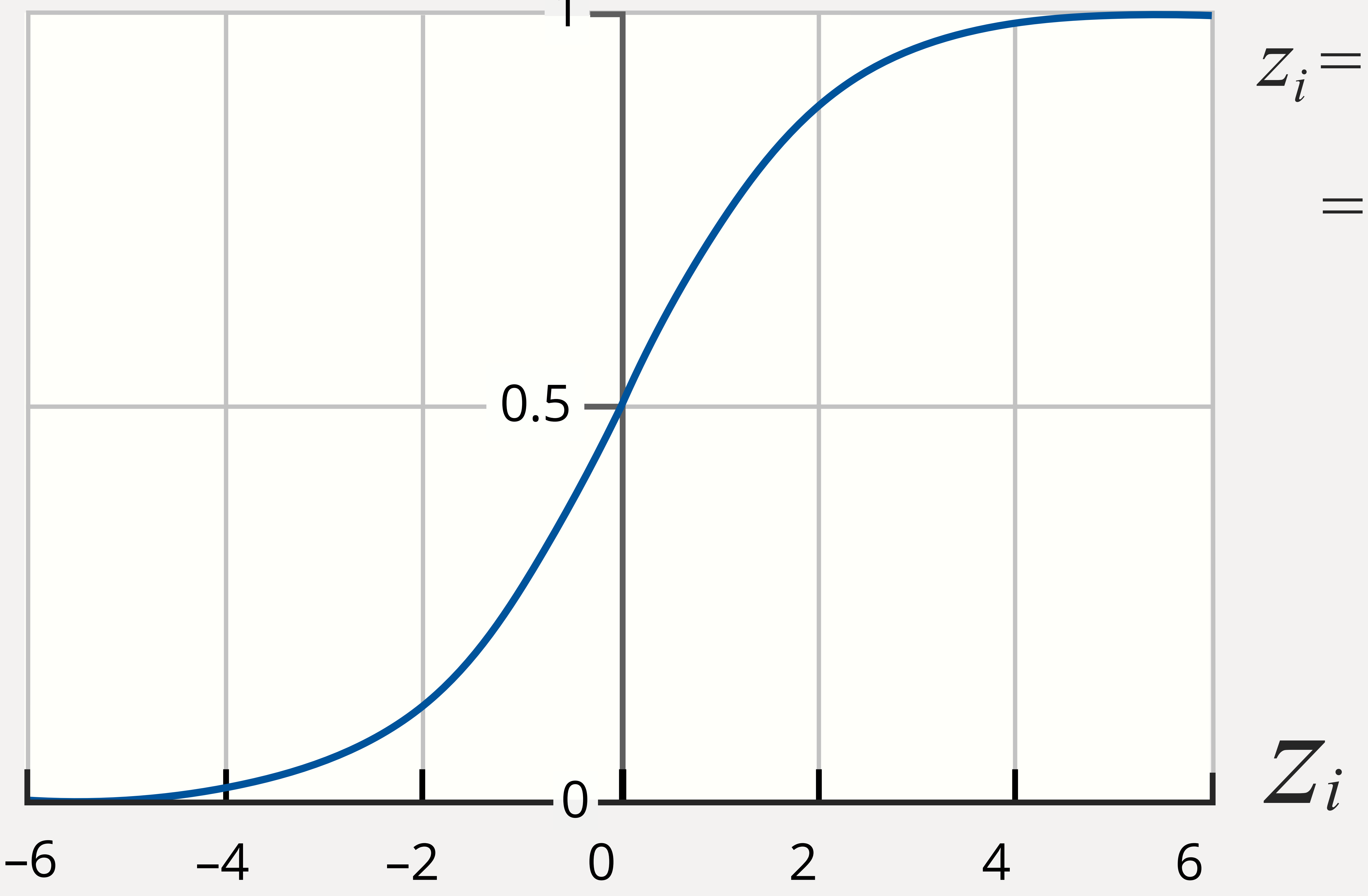
May be viewed as logistic regression on K latent features, rather than directly on the M components of raw data

Many Ways of Writing “4”



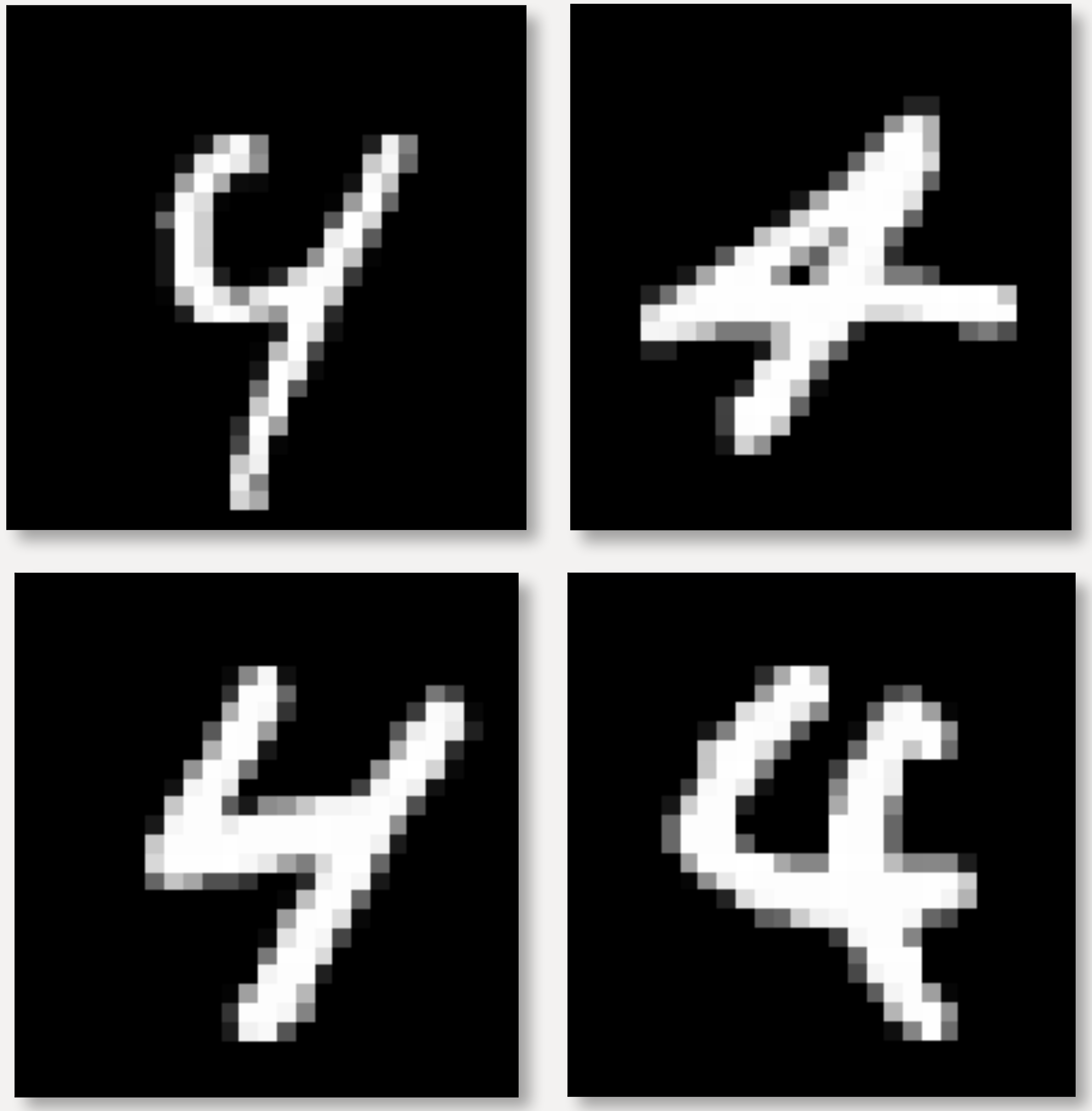
Logistic Regression

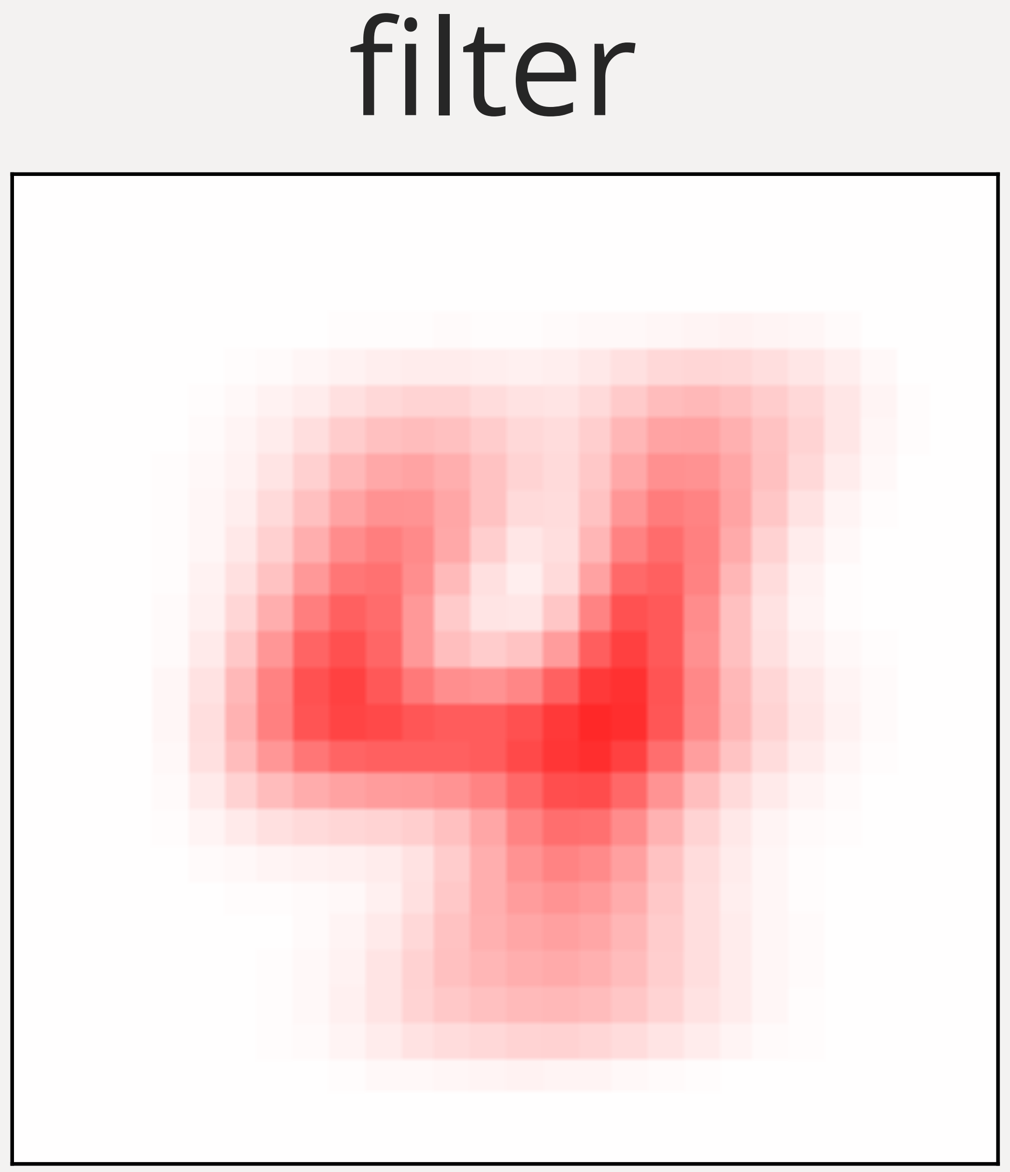
Sigmoid Function $p(y_i = 1 | x_i) = \sigma(z_i)$



$$z_i = (b_1 \times x_{i1}) + (b_2 \times x_{i2}) + \cdots + (b_M \times x_{iM}) + b_0$$

$$= b_0 + x_i \odot b$$



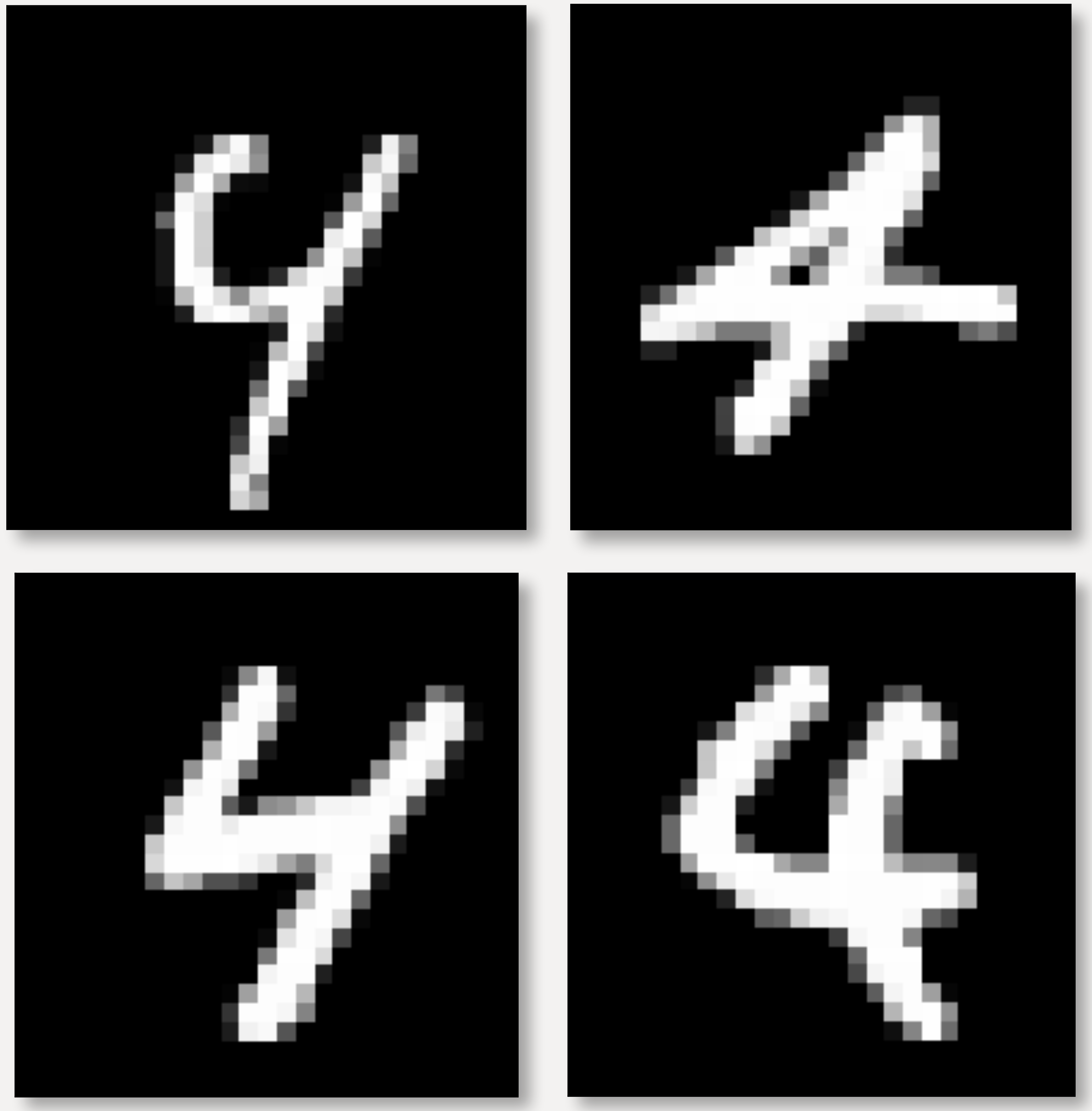


Single Filter (Shallow Learning)

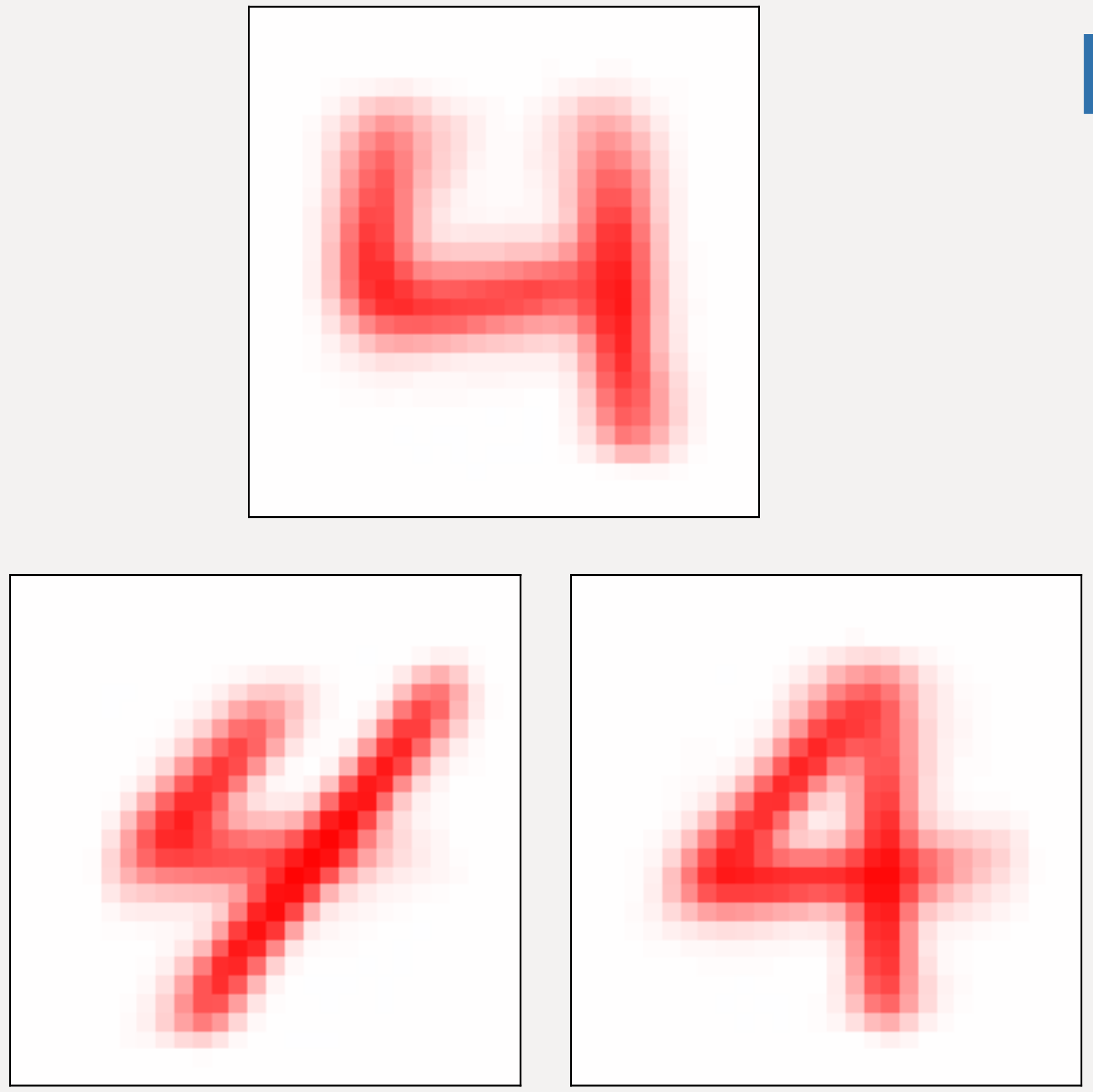
- Use of single filter only looks for the average shape

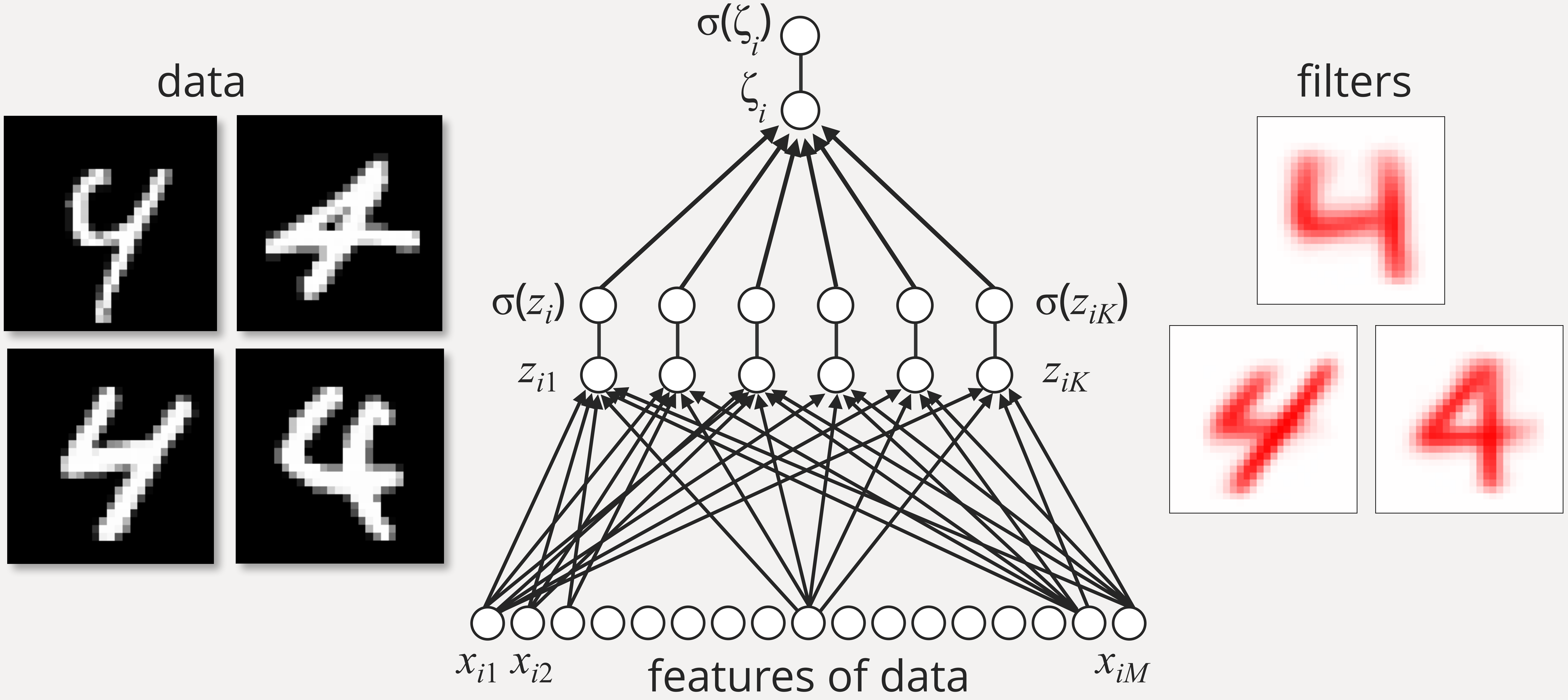
Multiple Filters

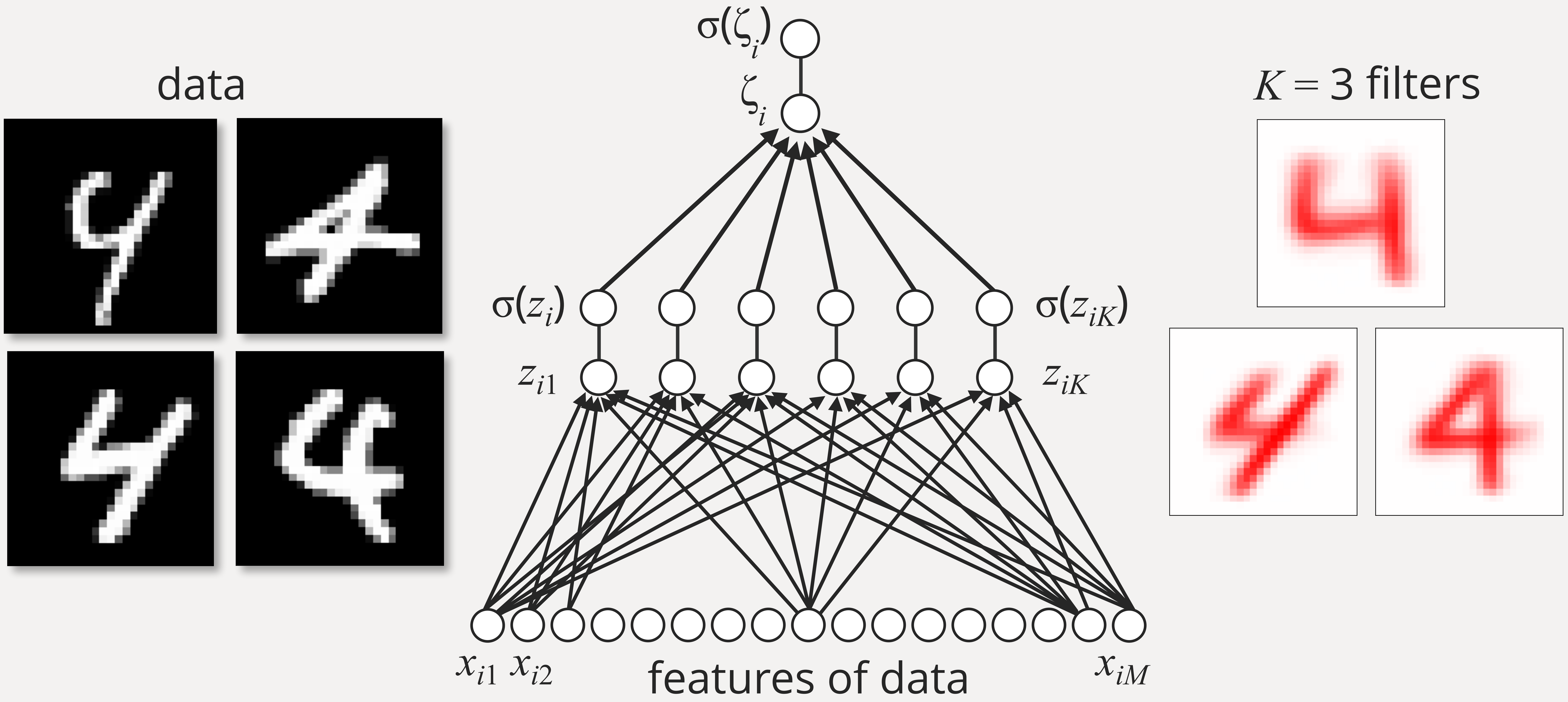
- Can look for **subtypes** indicative of different ways of writing “4”



filters




data
$\sigma(\zeta_i)$
ζ_i
$\sigma(z_i)$
$\sigma(z_{iK})$
z_{i1}
z_{iK}
x_{i1} x_{i2}
features of data
x_{iM}
filters



data
$\sigma(\zeta_i)$
ζ_i
$\sigma(z_i)$
$\sigma(z_{iK})$
z_{i1}
z_{iK}
x_{i1} x_{i2}
features of data
x_{iM}
$K = 3$ filters

Introducing K Filters

$$
\left.
\begin{aligned}
z_{i1} &= b_{01} + x_i \odot b_1 \\
z_{i2} &= b_{02} + x_i \odot b_2 \\
&\vdots \\
z_{iK} &= b_{0K} + x_i \odot b_K
\end{aligned}
\right]
$$

project data x_i onto
K filters: $b_1, \ldots, \mathrm{b}_K$