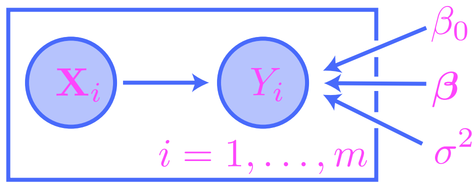
Problem 1: Suppose that we have three binary random variables X, W, and H, which indicate whether a person exercises regularly (X), whether they are overweight (W), and whether they have heart disease (H). Discuss several causal structure involving these variables.

Problem 2: Let X and Y be binary random variables that indicate whether a person has *hydromechanical trepidation syndrome* (X) and whether they test positive (Y) for it. Suppose that

$$p(y = 1|x = 1) = 0.98 \quad \text{and} \quad p(y = 1|x = 0) = 0.1.$$

Express the flow of information from X to Y as a stochastic link by explicitly writing down a link function.

Problem 3: Explicitly draw the full graphical structure for a plated linear regression model



when $m = 3$.

Problem 4: Given a linear regression model with predictors $\mathbf{X}$, response Y, and parameters β_0, $\boldsymbol{\beta}$, and σ^2, we saw in class that

$$R \mid \mathbf{X} = \mathbf{x} \sim N(0, \sigma^2)$$

where $R = Y - \beta_0 - \mathbf{X}^\intercal \boldsymbol{\beta}$ is the error term. Prove this claim.

Problem 5: Compute the inverse of the sigmoid function $\sigma(x) = 1/(1 + e^{-x})$. Interpret its meaning.

Problem 6: The *confusion matrix* for the logistic regression model discussed in class is given by

	$\hat{y}=1$	$\hat{y}=0$
$y=1$	417	95
$y=0$	90	422

where y is the true class and $\hat{y}$ the predicted class of an instance in the dataset. In this problem, a *positive* class label corresponds to 1, while a *negative* class label corresponds to 0.

(a) Compute the *accuracy* of the classifier, which is the proportion of correctly classified instances out of all total instances.

(b) Compute the *precision* of the classifier, which is the proportion of all true positive predictions out of all positive predictions. (High precision = avoids false positives.)

(c) Compute the *recall* (or *sensitivity*) of the classifier, which is the proportion of all true positive predictions out of all actual positive instances. (High sensitivity = avoids false negatives.)

(d) Discuss situations in which the primary interest is in classifiers with high precision, versus a situation in which the primary interest is high sensitivity.

Problem 7: Describe a possible neural network architecture for datasets of the form

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

where the y_i's are drawn from a continuous random variable Y.