Activity 1:
Understanding Logistic Regression

MMCi Applied DS

Why is this critical?

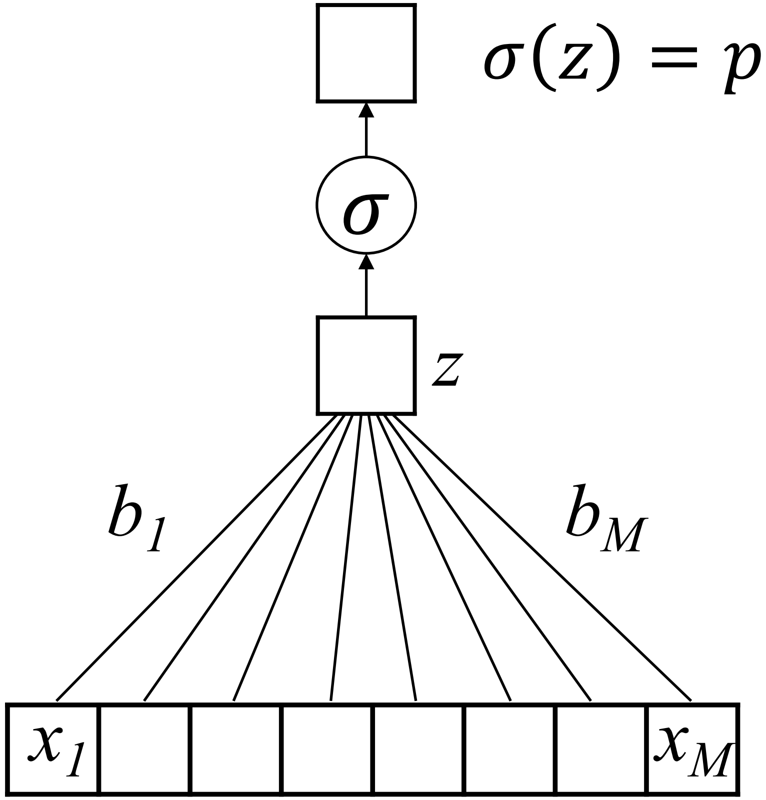
We need to understand (a) predictive modeling concepts, and (b) components found in all predictive models. Logistic regression is the simplest example.

Models are equations intended to capture relationships between real-world quantities. *Example*: $F = ma$

All predictive models have:

- Input *features* (x)
- A predicted *label*, or *value* (y)
- Numeric *parameters* that can be adjusted, or *learned*, to more accurately reflect the observed relationship between x and y in a training dataset

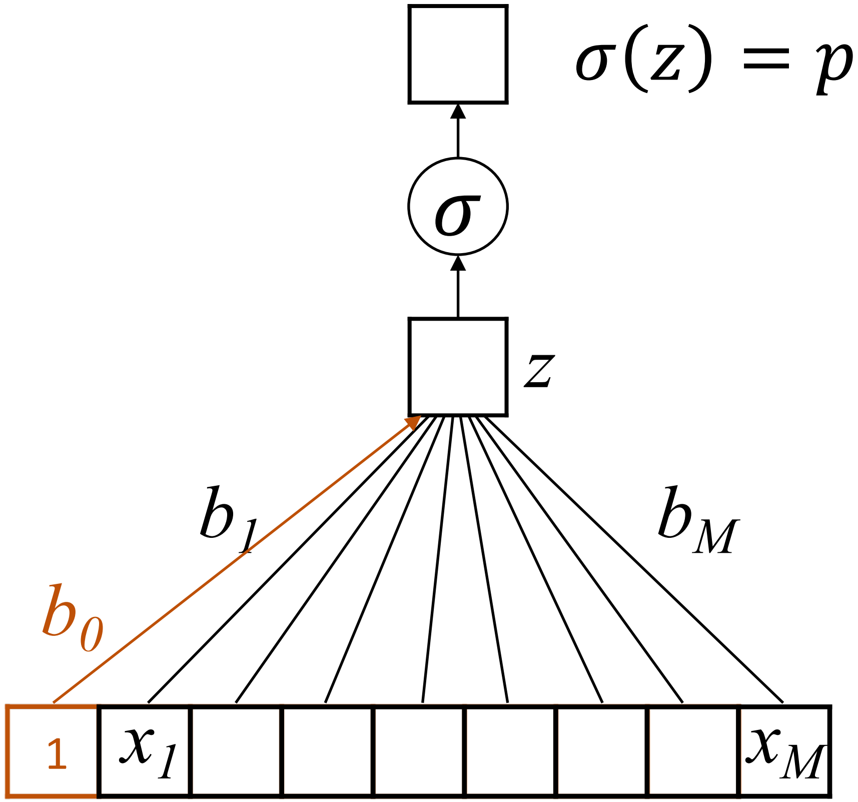
<u>Logistic Regression</u>: a linear model with a logistic “link” function that converts the prediction to a probability



$$\sigma(z) = p(y = 1|x)$$

$$p(y = 1|x) = \sigma(b_1 x_1 + b_2 x_2 + \cdots + b_M x_M)$$

Logistic Regression: a linear model with a logistic "link" function that converts the prediction to a probability



$\sigma(z) = p(y = 1|x)$

A way to represent the *intercept* (i.e., *bias*) term

$$p(y = 1|x) = \sigma(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_M x_M)$$

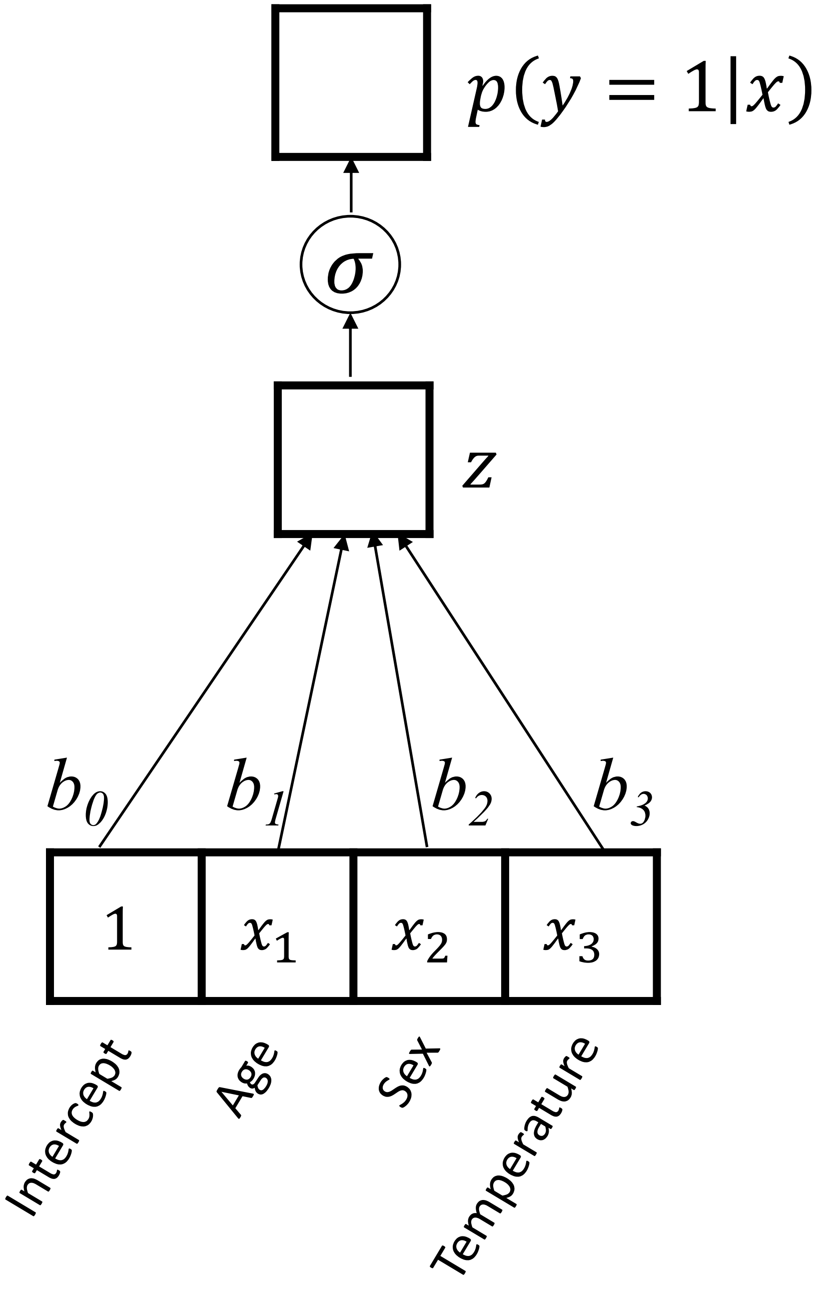
Part I. Suppose you have a previously trained logistic regression model that predicts a patient's probability of dying during their ICU stay based on their age, sex, and temperature.

1. What are the *features* (i.e., *predictors*)?

2. What is the associated *label* (y), and how does it relate to p in the diagram at right?

3. Which values in the diagram would we *learn* if we were to train this model, and what are they called?

4. Write the equation corresponding to this model. You may write $\sigma(z)$ (or sigma(z)) to denote the *logistic* (i.e., *sigmoid*) function applied to z.



Part II. A 70-year old ($x_1 = 70$) woman (female sex; $x_2 = 1$) comes into the ICU. Her temperature is 39 degrees Celsius ($x_3 = 39$). Looking closely at your model, you find that the previously learned values of b are as follows:

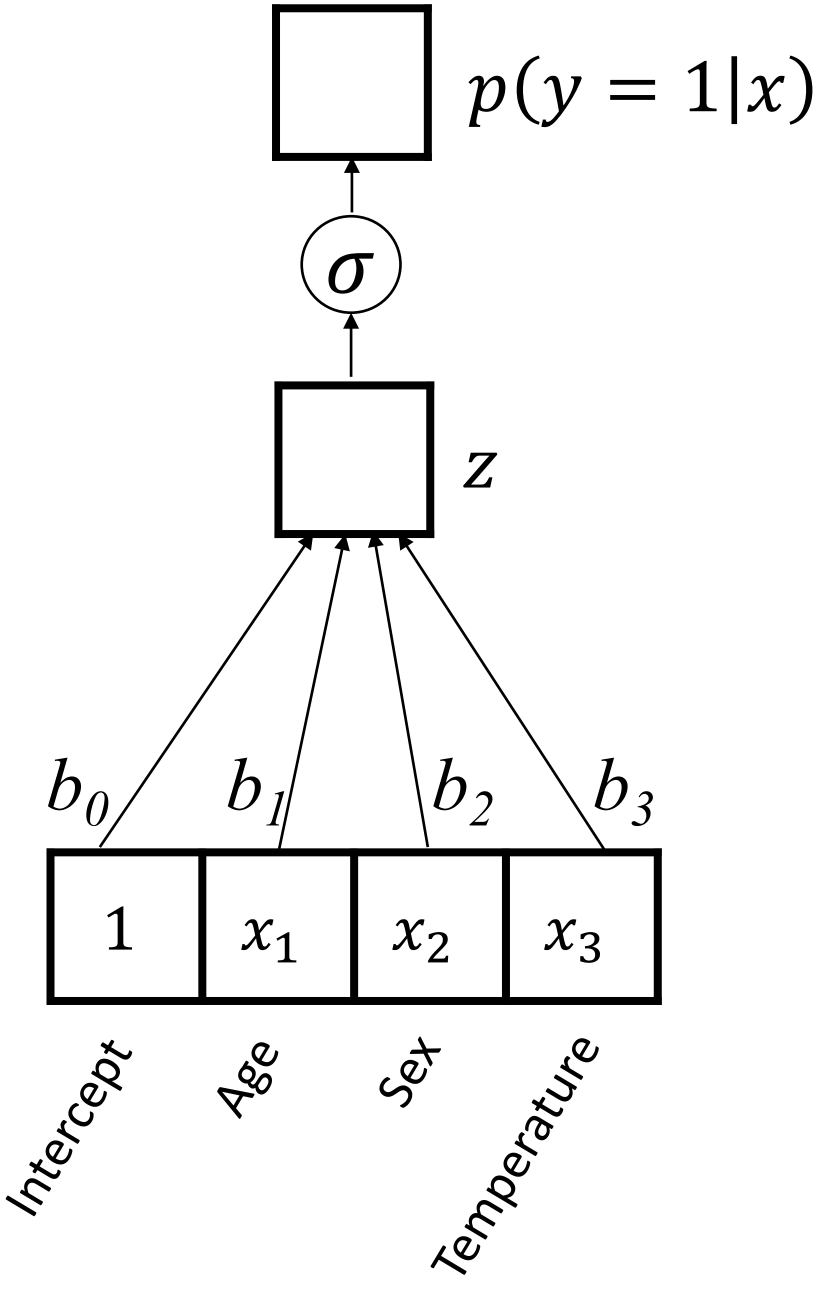
$$b_0 = -20 \qquad b_1 = .1 \qquad b_2 = -.5 \qquad b_3 = .3$$

1. Calculate the model-predicted <u>log-odds</u> (z) that this patient will die during her ICU stay.

2. Calculate and interpret the model-predicted <u>odds</u> (e^z) that this patient will die during her ICU stay.

3. Calculate and interpret the model-predicted <u>probability</u> ($\frac{e^z}{1+e^z}$) that this patient will die during her ICU stay.

4. Repeat these calculations for a 70-year old *man* with a 39-degree Celsius temperature.



Part III. Unfortunately, your patient's condition deteriorates, and despite everyone's best efforts, she passes away in the ICU.

Consider the relationship between this outcome and the probability of mortality predicted by your model.

The following questions do not have clear yes or no answers. They are designed to get you thinking about model fit and performance.

1. Was your model's prediction *correct*? Why or why not?
2. Was your model's prediction *good*? Why or why not?
3. Is there any additional information you could collect to help you answer (1) and (2)?

Next course weekend, we will continue to explore these questions.