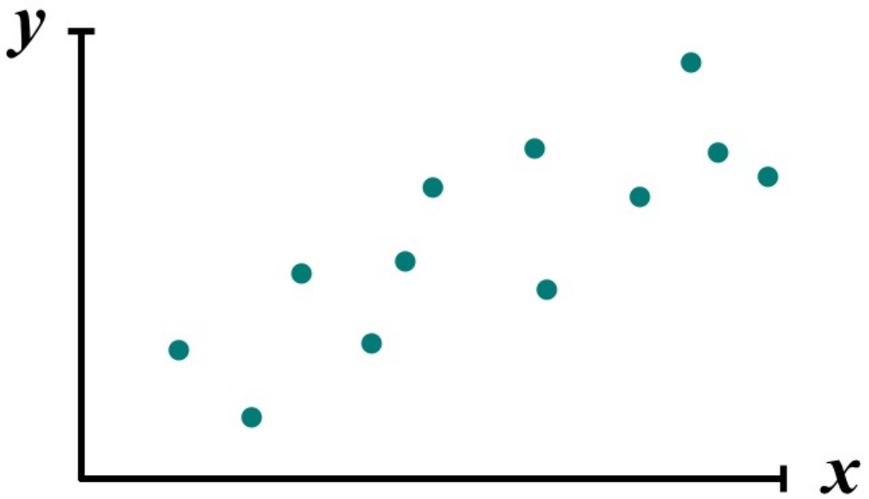
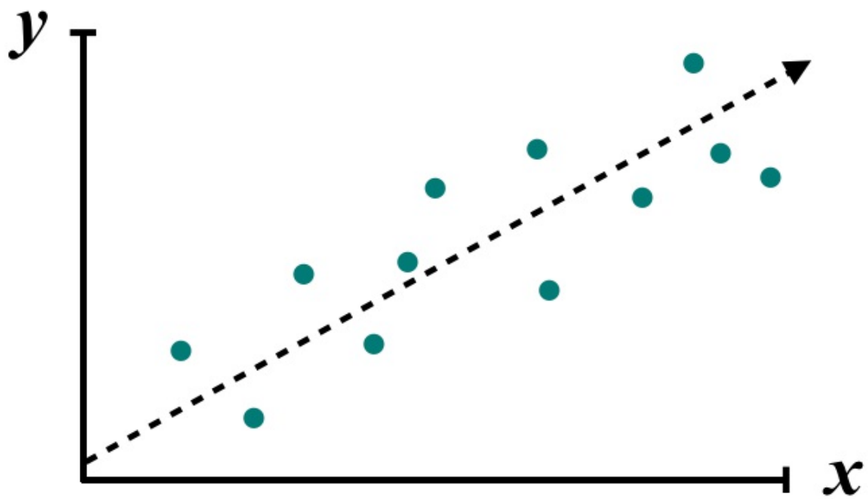
SUPERVISED LEARNING IN R: CLASSIFICATION

Making binary predictions with regression

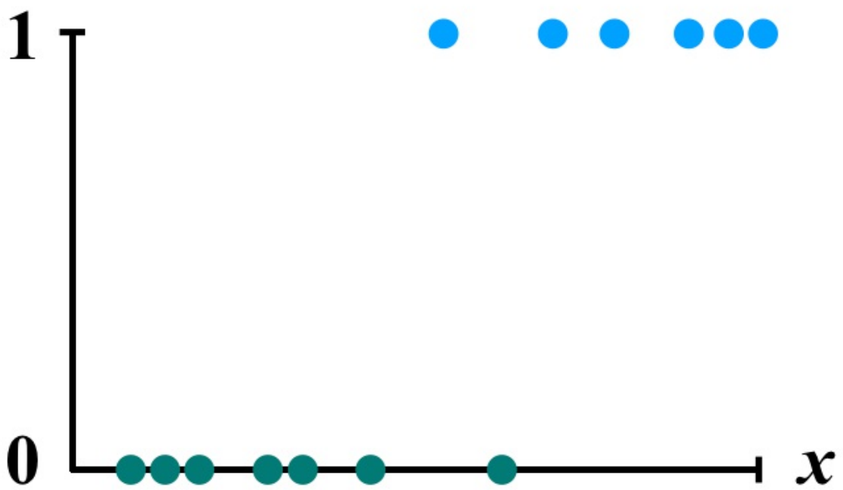
Brett Lantz
Instructor

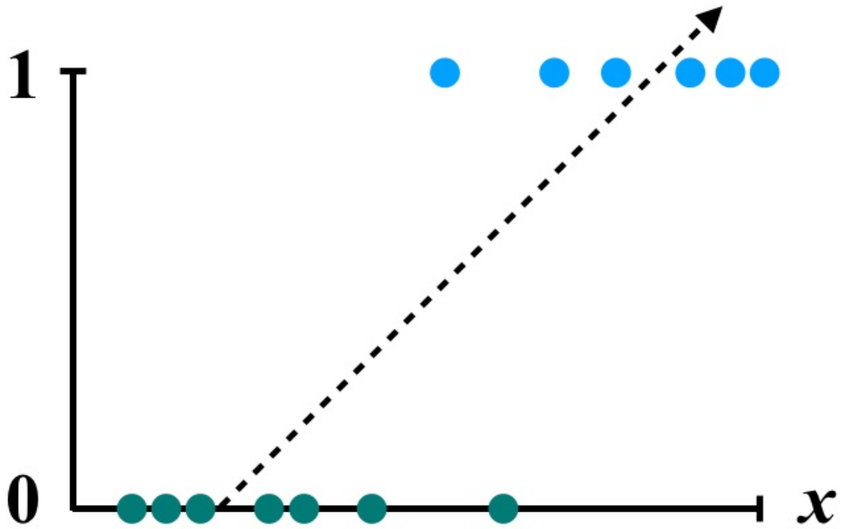
Introducing linear regression



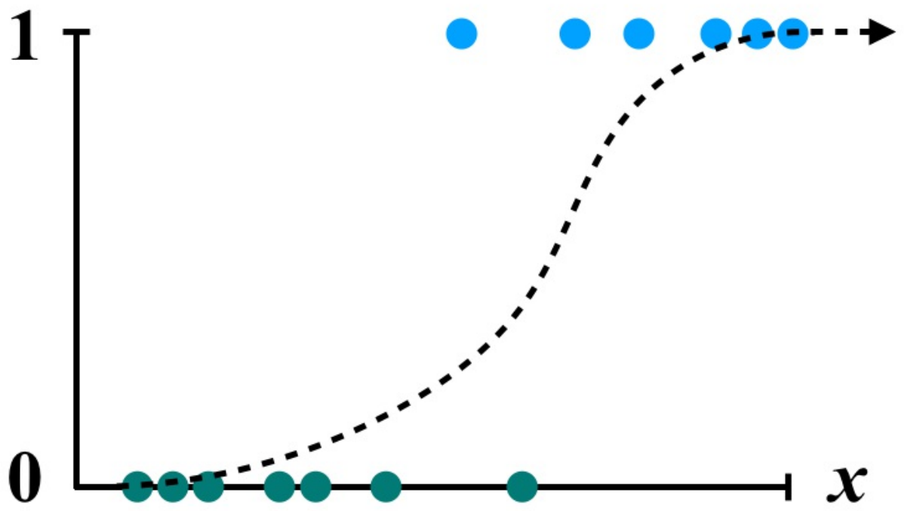


Regression for binary classification

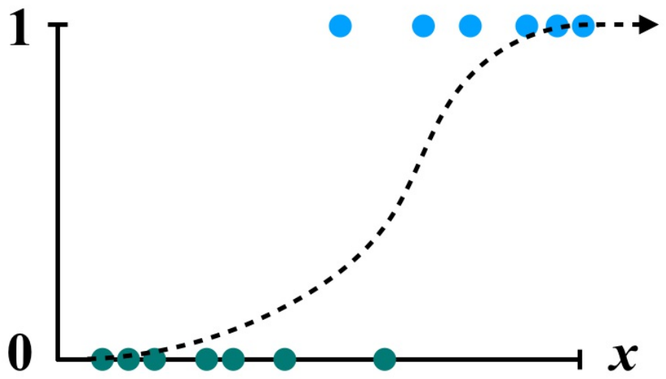




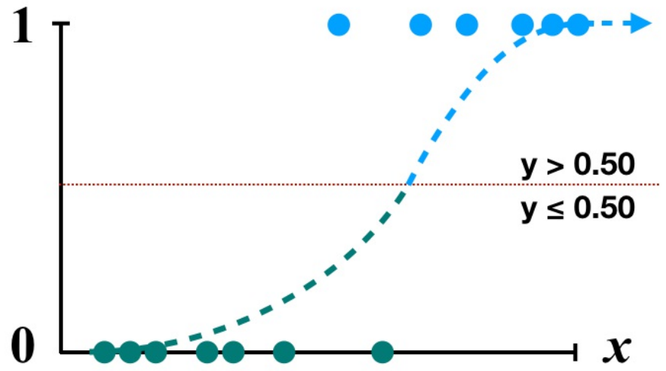
Introducing logistic regression



Making predictions with logistic regression




```
m <- glm(y ~ x1 + x2 + x3,
         data = my_dataset,
         family = "binomial")
```




```
prob <- predict(m, test_dataset,
          type = "response")
```

```
pred <- ifelse(prob > 0.50, 1, 0)
```

SUPERVISED LEARNING IN R: CLASSIFICATION

Let's practice!

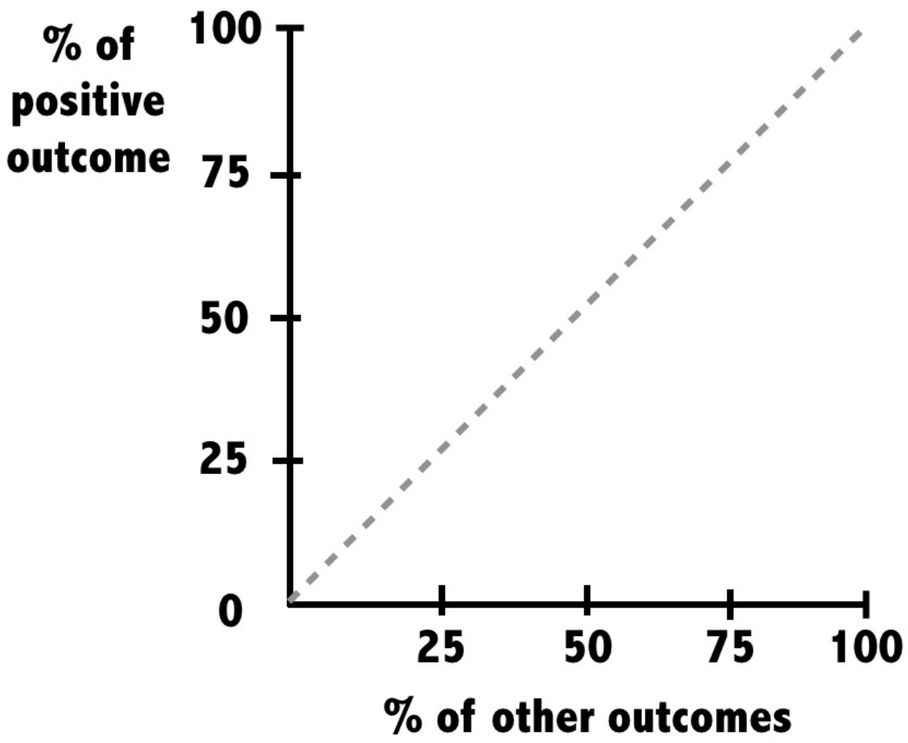
SUPERVISED LEARNING IN R: CLASSIFICATION

Model performance tradeoffs

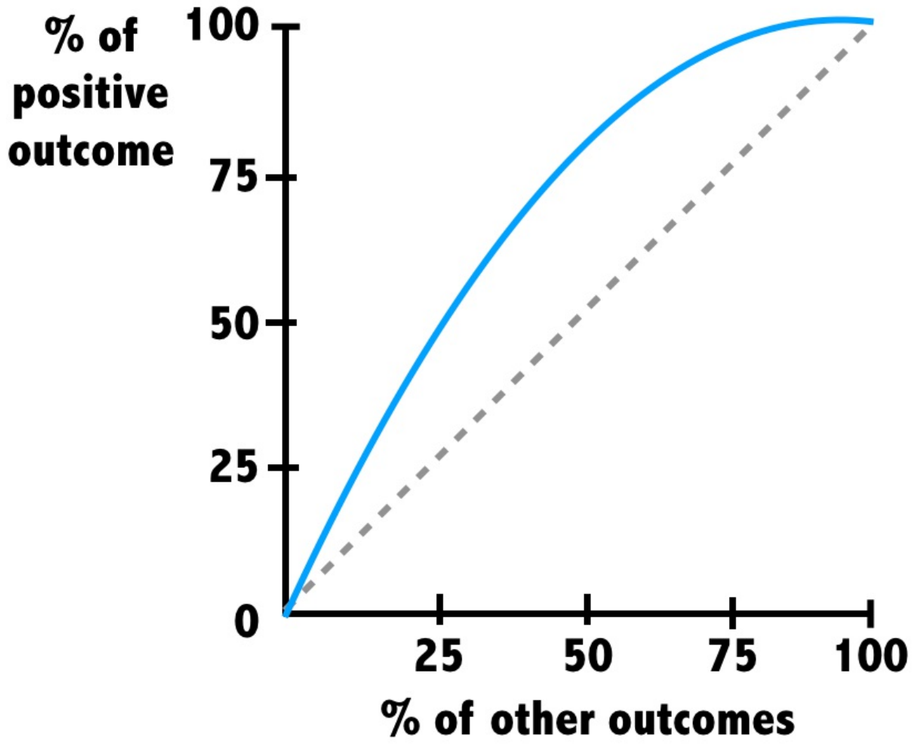
Brett Lantz
Instructor

Understanding ROC curves

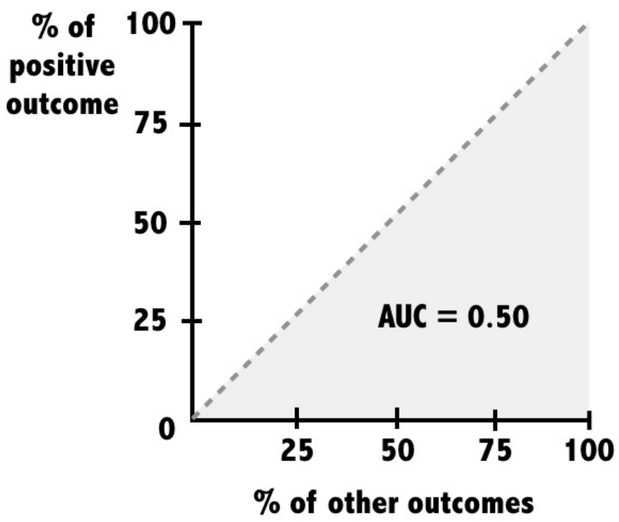
XOXOXOXOXOXOXOXOXOXOXO

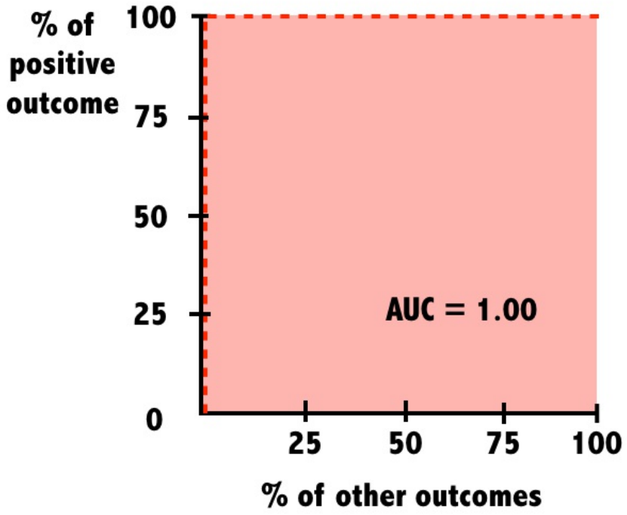


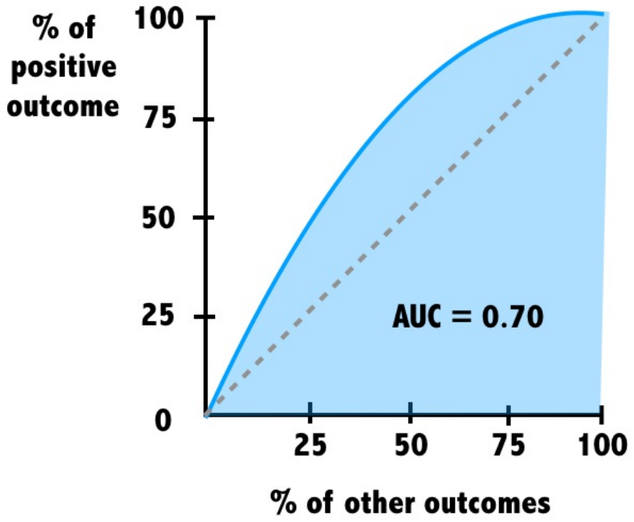
XXOXXXXXOXXOOXOXOOOOO



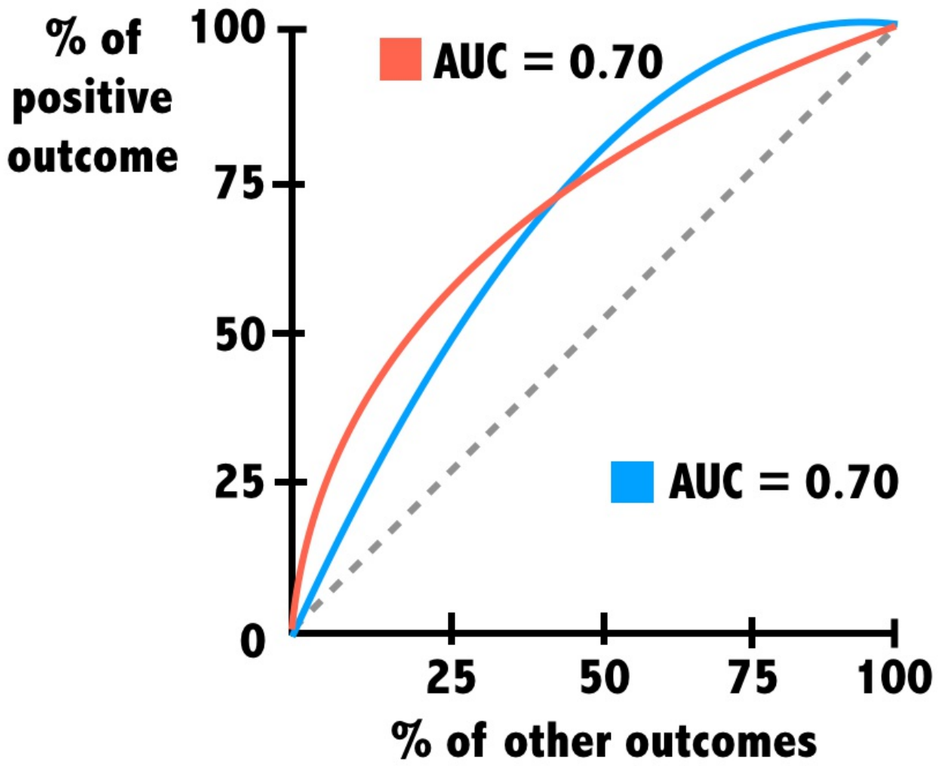
Area under the ROC curve







Using AUC and ROC appropriately



SUPERVISED LEARNING IN R: CLASSIFICATION

Let's practice!

SUPERVISED LEARNING IN R: CLASSIFICATION

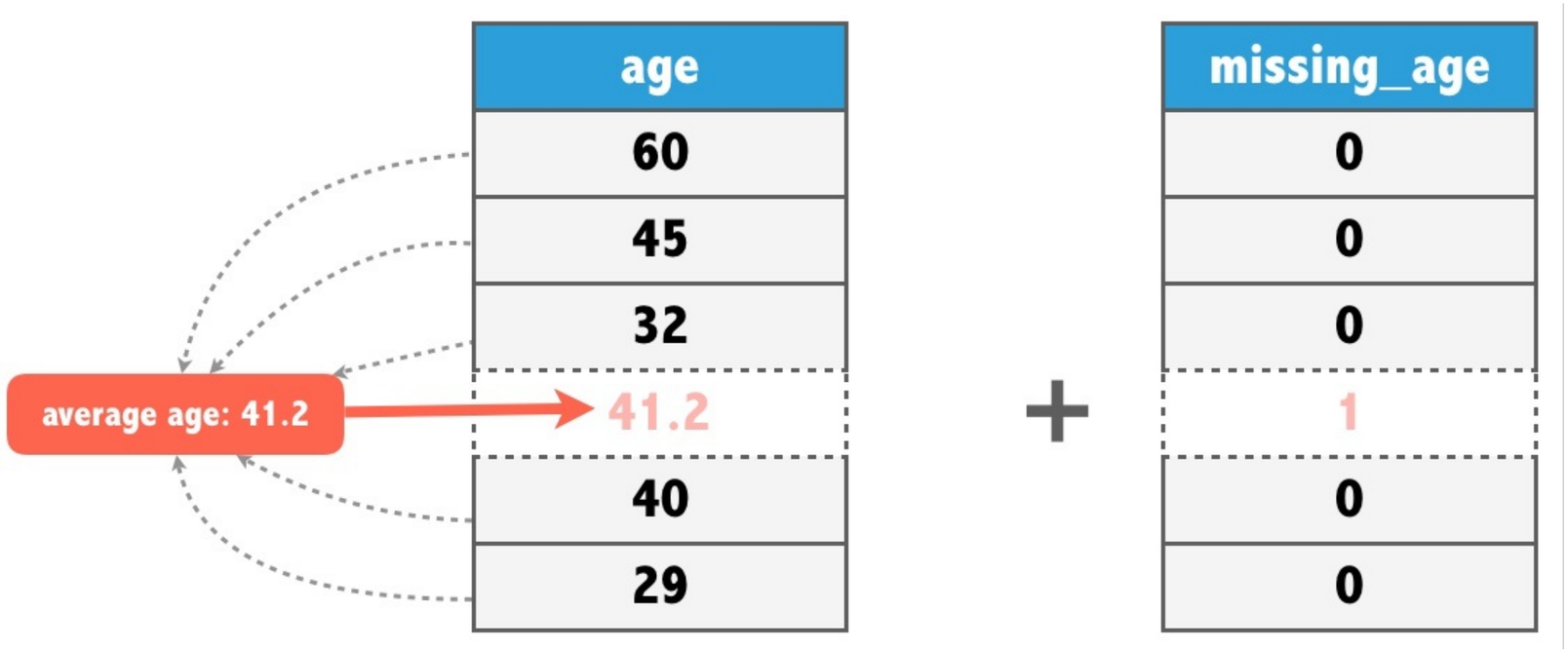
Dummy variables, missing data, and interactions

Brett Lantz
Instructor

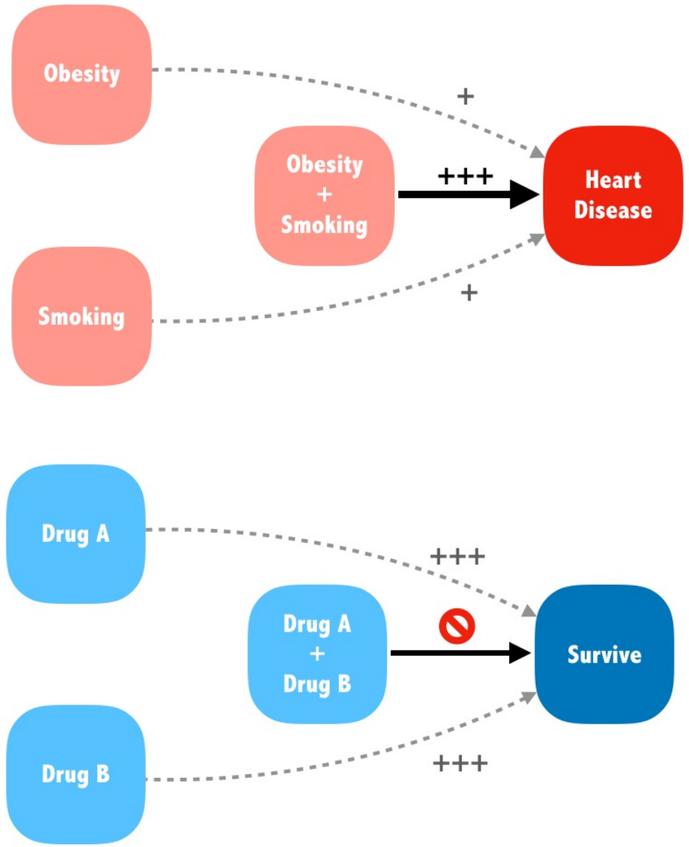
Dummy coding categorical data

```
# create gender factor
my_data$gender <- factor(my_data$gender,
                         levels = c(0, 1, 2),
                         labels = c("Male", "Female", "Other"))
```

Imputing missing data



Interaction effects




```
# interaction of obesity and smoking
glm(disease ~ obesity * smoking,
      data = health,
      family = "binomial")
```

SUPERVISED LEARNING IN R: CLASSIFICATION

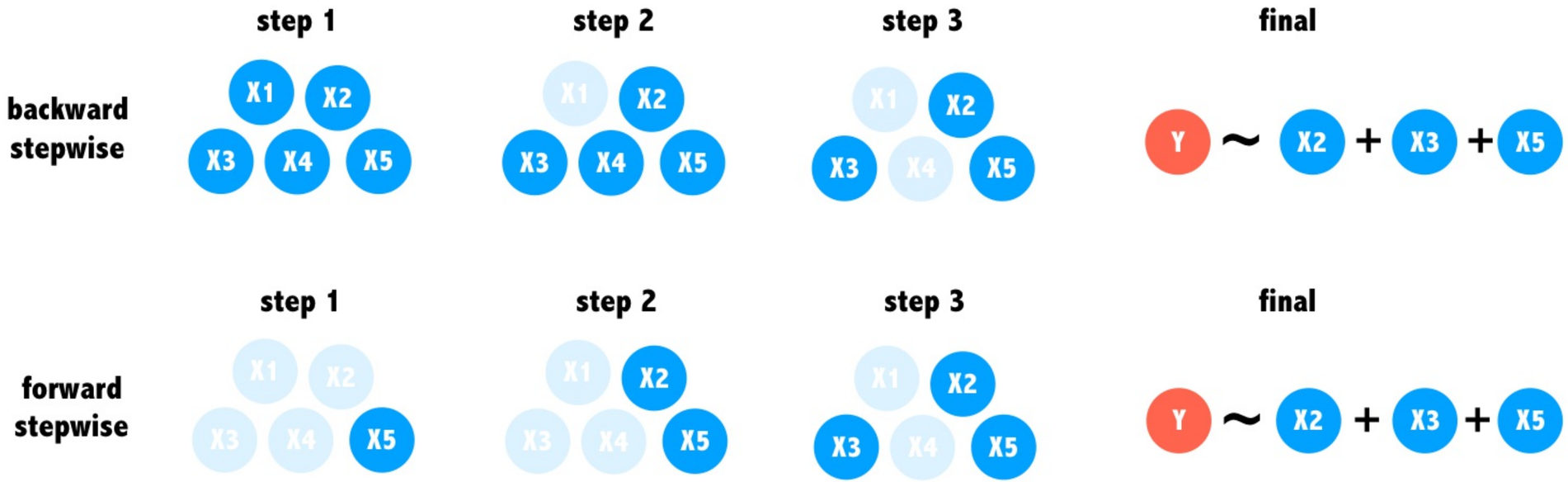
Let's practice!

SUPERVISED LEARNING IN R: CLASSIFICATION

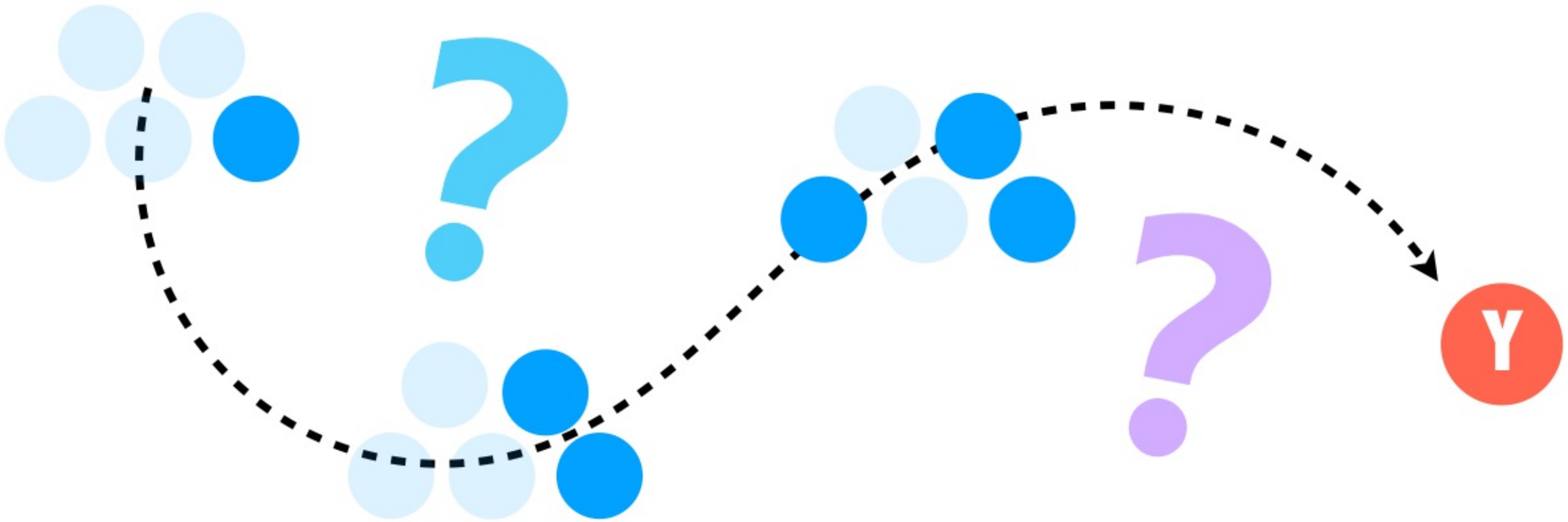
Automatic feature selection

Brett Lantz
Instructor

Stepwise regression



Stepwise regression caveats



SUPERVISED LEARNING IN R: CLASSIFICATION

Let's practice!