Lab 10

Not graded, just practice

Katie Schuler

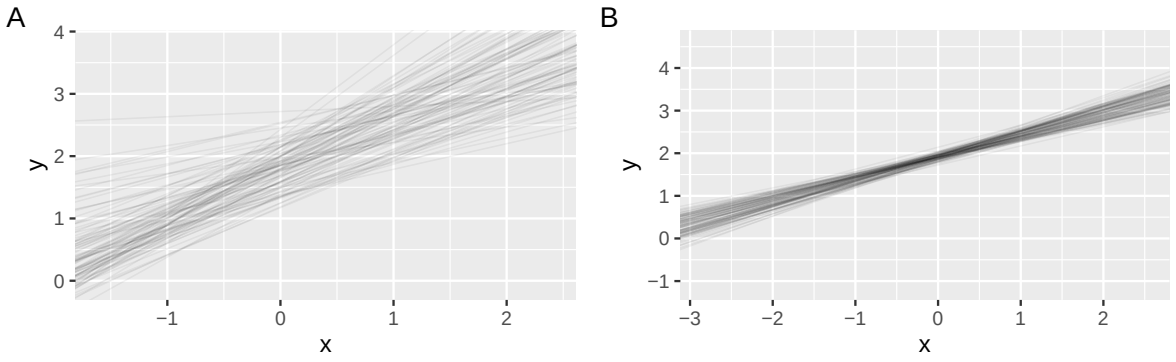
2024-11-14

1 Model reliability

a. As we collect more data, our parameter estimates

- (A) become more reliable
- (B) become less reliable
- (C) stay the same

b. Each figure below plots 100 bootstrapped models with data drawn from the same population. In one figure, the model is fit to 10 data points. In the other, each model is fit to 200 data points. Which figure shows the model fit to 200 data points?



- (A) Figure A
- (B) Figure B

- (C) Both figures have the same number of data points

c. As we collect more data, what happens to the confidence interval around our parameter estimates?

- (A) It gets bigger (wider)
- (B) It stays the same
- (C) It gets smaller (narrower)

d. True or false, we can obtain confidence intervals around parameter estimates for models in the same we we did for point estimates like the mean.

- (A) True
- (B) False

e. Model reliability asks how certain we can be about our parameter estimates. Why is there uncertainty around our parameter estimates?

```
"Because we are interested in the model parameters that
best describe the population from which the sample was
 drawn. Due to sampling error, we can expect some
 variability in the model parameters."
```

f. The figure below shows the model fit for a sample of 10 participants. Suppose we repeated our experiment with 10 new participants. True or false, fitting the same model to these new data would yield approximately the same parameter estimates.

- (A) True
- (B) False

g. True or false, a model with high accuracy must also have high reliability.

- (A) True
- (B) False

2 Classification

a. Which of the following aspects of model building apply to classification models? (Choose all that apply)

- (A) specify
- (B) fit
- (C) accuracy
- (D) reliability

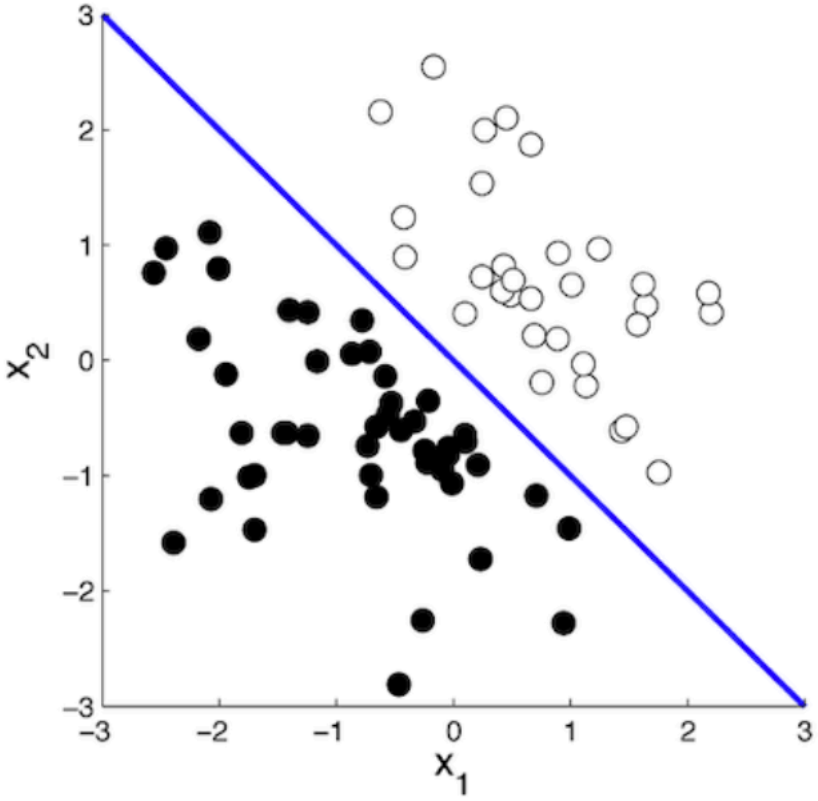
b. What is the difference between regression and classification?

```
"Regression predicts a continuous response varaible,
classification predicts a discrete response variable"
```

c. Which accuracy metric is best applied to classification models?

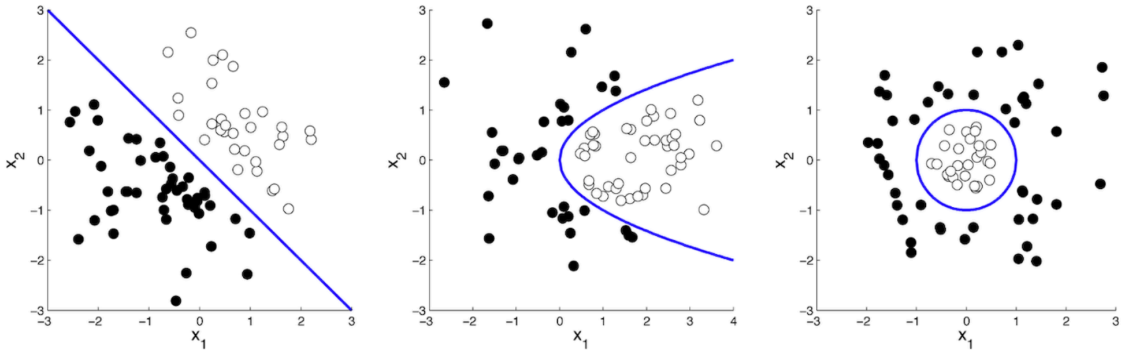
- (A) R squared
- (B) RMSE
- (C) Pecent correct
- (D) All of the above

d. In the figure below, which aspect shows the response variable?



- (A) x-axis
- (B) y-axis
- (C) color of points
- (D) blue line

e. Which figure below could show a plotted classification model? Choose all that apply.



- (A) Figure A
- (B) Figure B
- (C) Figure C

f. Name two kinds of linear classifiers

```
"Any 2 of those mentioned in class:
Logistic regression
Linear discriminant analysis (LDA)
Linear support vector machines (SVM)
Nearest-prototype classifiers
Naive Bayes classifiers
"
```

4 Classification in R

a. We can impliment classification via

- (A) linear regression
- (B) cubic polynomials
- (C) logistic regression
- (D) nearest-neighbor regression
- (E) all of the above

b. True or false, in R, we can perform logistic regression with a generalized linear model.

- (A) True
- (B) False

c. What 3 elements do all GLMs have?

```
"1. A particular distribution for modeling the response variable
2. A linear model
3. A link function
"
```

d. What is the link function for logistic regression?

- (A) logistic function
- (B) polynomial expansion
- (C) log transformation
- (D) inverse transformation

f. Which of the following fits a logistic regression model in R? Choose one.

```
# code A
glm(y ~ x, data = data, family = "binomial")

# code B
data %>%
    specify(y ~ x) %>%
    fit()

# code C
logistic_reg %>%
    set_engine("glm") %>%
    fit(y ~ x, data = data)
```

- (A) code A
- (B) code B
- (C) code C
- (D) code A and B only
- (E) all of the above