# STAT 215A Fall 2019 Week 13

Tiffany Tang

11/22/19

# Plan for Today

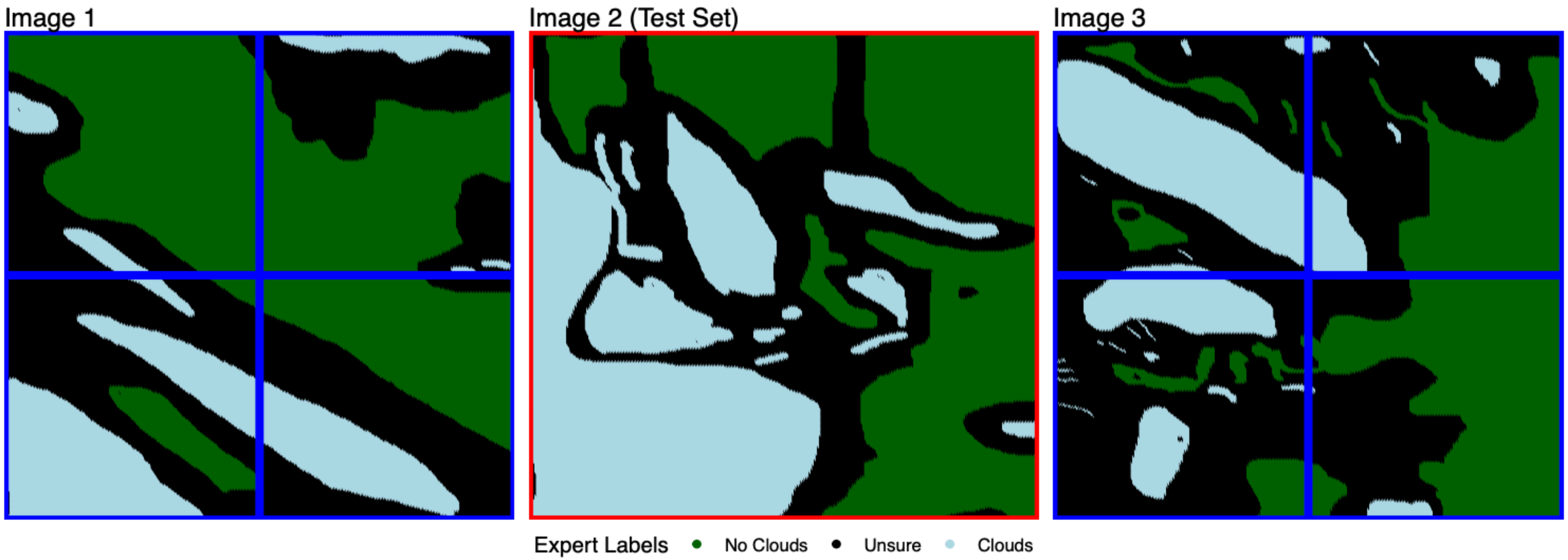
- Post lab 4 discussion
  - Data splitting
  - Generalizability
- ROC/PR Curves and AUC
- High-level overview of data science life cycle
- Introduction to final project
- Next time: Rebecca Barter guest lecture; research talk

# Data Splitting

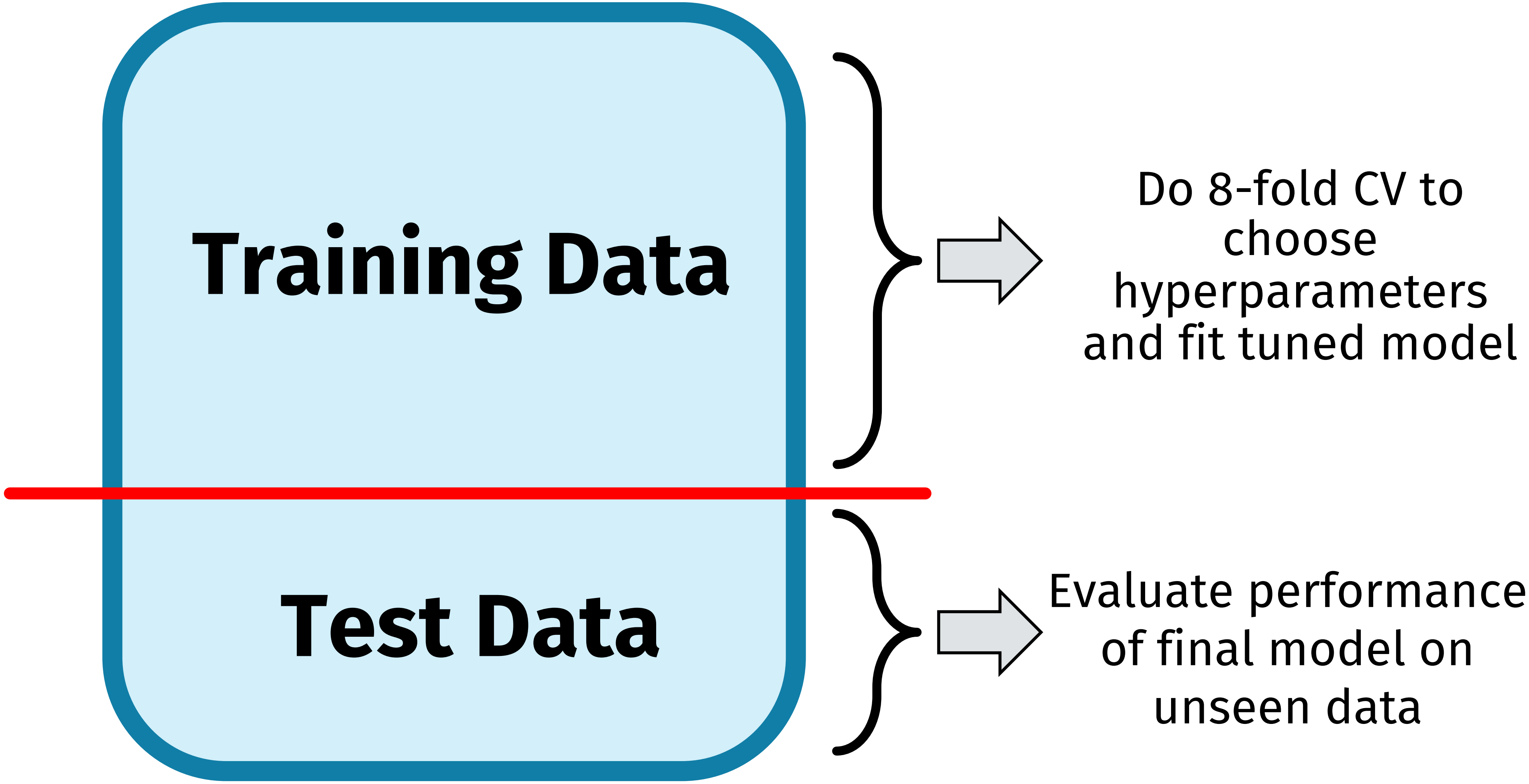
- **Goal:** use data splitting to obtain an unbiased estimate for the prediction error on *future* data
- If we ignore spatial dependencies and randomly split pixels into training/test, we will obtain an upwardly biased estimate of the test error
- **Solution:** mimic the process of obtaining new data

# Data Splitting

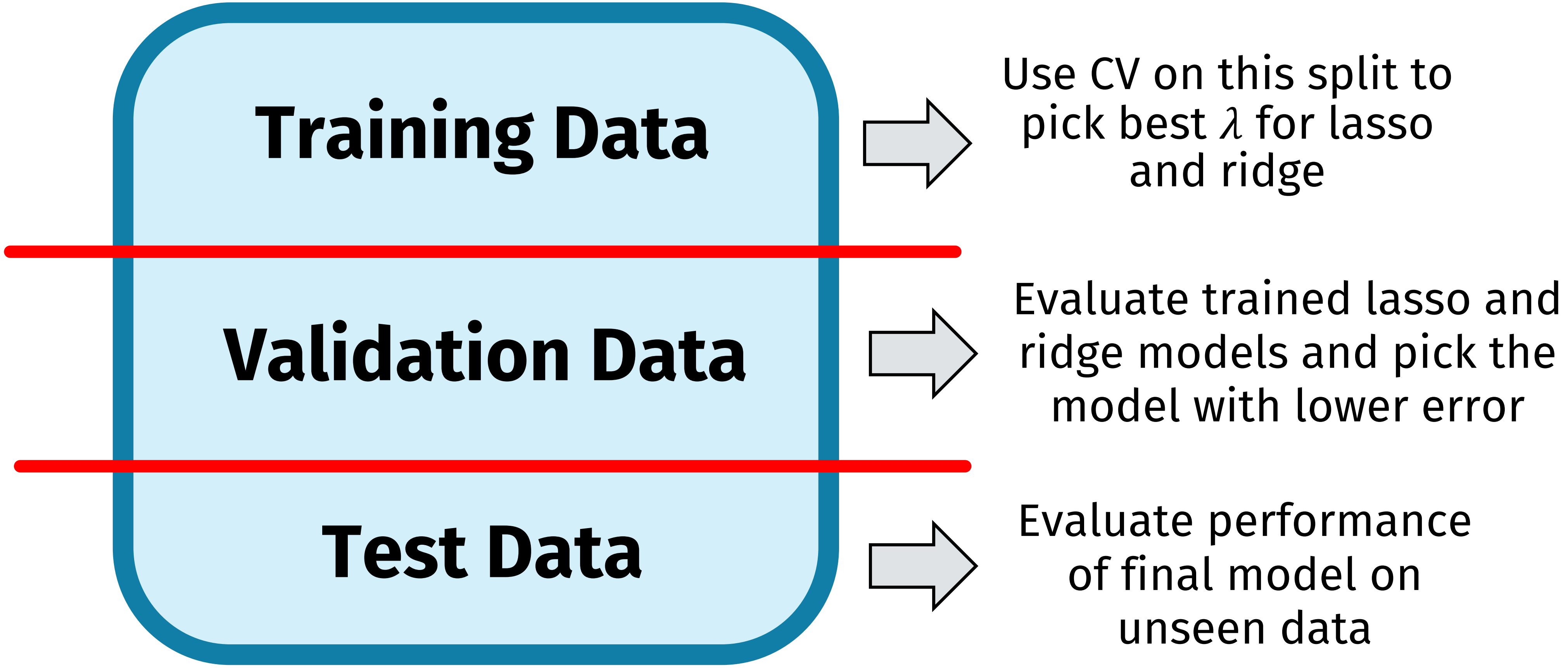
- **Goal:** use data splitting to obtain an unbiased estimate for the prediction error on *future* data
- If we ignore spatial dependencies and randomly split pixels into training/test, we will obtain an upwardly biased estimate of the test error
- **Solution:** mimic the process of obtaining new data



# Data Splitting + Tuning Hyperparameters



# Data Splitting + Tuning Hyperparameters + Multiple Methods




- *Can* repeat this data splitting B times to get a variance estimate of the test error
- This gives you an unbiased estimate of the prediction error for the **statistical learning process**, NOT a specific model

# Evaluation Metrics for Classification

# How to evaluate your classification methods

- Going beyond classification error…
- What if we have class imbalance?
  - Ex. sample of 100 people and only 10 have the disease → always predict healthy and get 90% classification accuracy!!

# Confusion Matrix

| Hypothesized class | True class: p | True class: n |
|---|---|---|
| Y | True Positives | False Positives |
| N | False Negatives | True Negatives |
| Column totals: | P | N |

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

$$\text{precision} = \frac{TP}{TP+FP} \qquad \text{recall} = \frac{TP}{P}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$

Fig. 1. Confusion matrix and common performance metrics calculated from it.

Source: Fawcett (2005)

# Confusion Matrix

| | | True class | |
|---|---|---|---|
| | | p | n |
| Hypothesized class | Y | True Positives | False Positives |
| | N | False Negatives | True Negatives |
| Column totals: | | P | N |

ROC curve

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

$$\text{precision} = \frac{TP}{TP+FP} \qquad \text{recall} = \frac{TP}{P}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$

Fig. 1. Confusion matrix and common performance metrics calculated from it.

Source: Fawcett (2005)

# Confusion Matrix

| Hypothesized class | | True class: p | True class: n |
|---|---|---|---|
| | **Y** | True Positives | False Positives |
| | **N** | False Negatives | True Negatives |
| **Column totals:** | | P | N |

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

precision-recall curve

$$\boxed{\text{precision} = \frac{TP}{TP+FP} \qquad \text{recall} = \frac{TP}{P}}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$

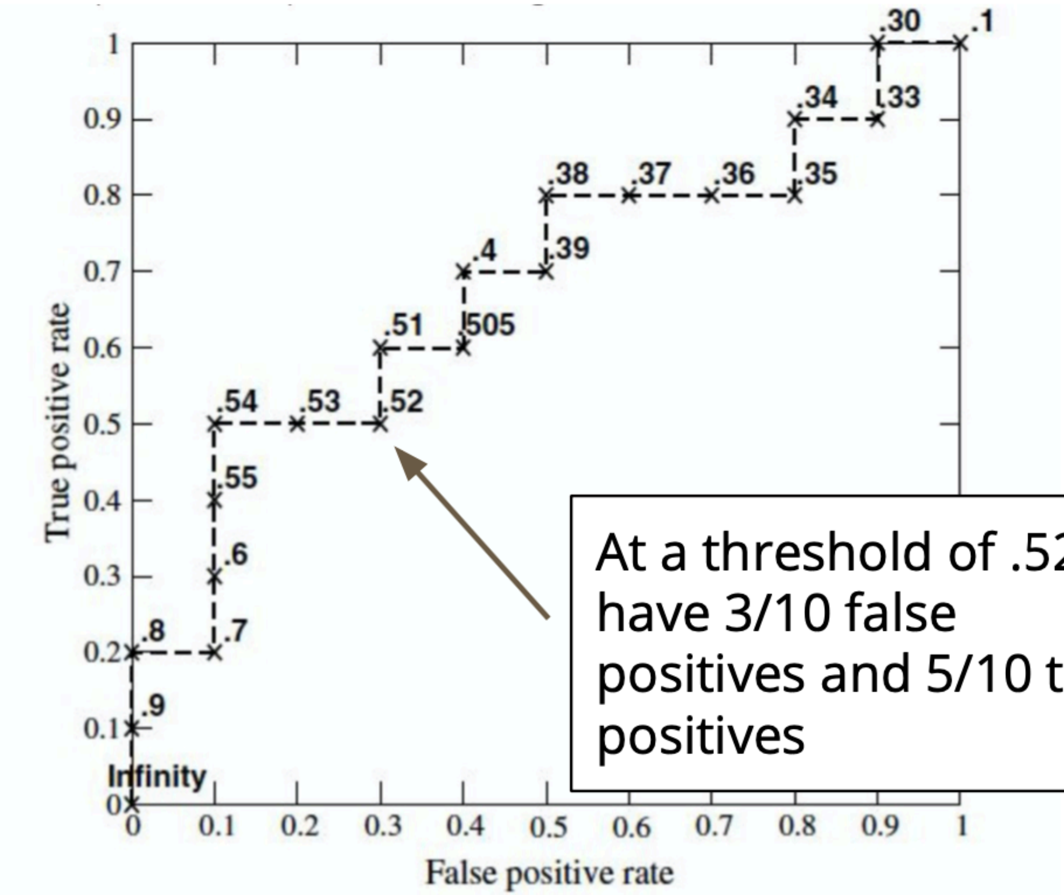
Fig. 1. Confusion matrix and common performance metrics calculated from it.

Source: Fawcett (2005)

# Receiver operating characteristics (ROC) curve

We can generate an ROC curve when the output of a classifier is a probability and we must choose a threshold for the final predicted class

| Inst# | Class | Score | Inst# | Class | Score |
|---|---|---|---|---|---|
| 1 | p | .9 | 11 | p | .4 |
| 2 | p | .8 | 12 | n | .39 |
| 3 | n | .7 | 13 | p | .38 |
| 4 | p | .6 | 14 | n | .37 |
| 5 | p | .55 | 15 | n | .36 |
| 6 | p | .54 | 16 | n | .35 |
| 7 | n | .53 | 17 | p | .34 |
| 8 | n | .52 | 18 | n | .33 |
| 9 | p | .51 | 19 | p | .30 |
| 10 | n | .505 | 20 | n | .1 |



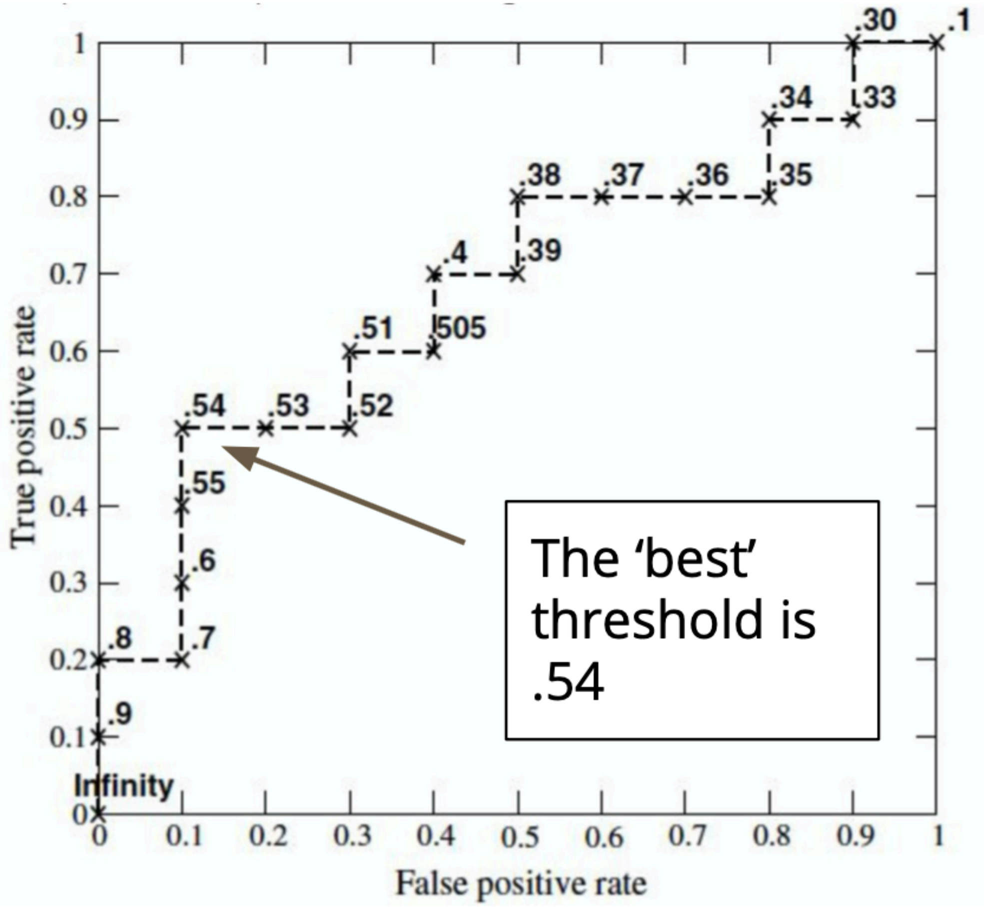

At a threshold of .52 we have 3/10 false positives and 5/10 true positives

Source: Fawcett (2005)

# Receiver operating characteristics (ROC) curve

We can generate an ROC curve when the output of a classifier is a probability and we must choose a threshold for the final predicted class

| Inst# | Class | Score | Inst# | Class | Score |
|---|---|---|---|---|---|
| 1 | p | .9 | 11 | p | .4 |
| 2 | p | .8 | 12 | n | .39 |
| 3 | n | .7 | 13 | p | .38 |
| 4 | p | .6 | 14 | n | .37 |
| 5 | p | .55 | 15 | n | .36 |
| 6 | p | .54 | 16 | n | .35 |
| 7 | n | .53 | 17 | p | .34 |
| 8 | n | .52 | 18 | n | .33 |
| 9 | p | .51 | 19 | p | .30 |
| 10 | n | .505 | 20 | n | .1 |




Source: Fawcett (2005)

# Area under the curve

The area under the curve (AUC) is a method for comparing algorithms and evaluating classifiers.

The AUC has an important statistical property:

> *The AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance*

Source: Fawcett (2005)

# Area under the curve

Care should be taken when using ROC curves to compare classifiers

- ❑ The ROC graph is often used to select the best classifiers simply by graphing them in ROC space and seeing which one dominates.
- ❑ This is misleading: it is analogous to taking the maximum of a set of accuracy figures from a single test set.
- ❑ Without a measure of **variance** we cannot compare classifiers

It is a good idea to the average of multiple ROC curves (e.g. via cross validation)

See Fawcett (2005) for examples on how to average

Source: Fawcett (2005)

# ROC vs Precision-Recall (PR) Curves

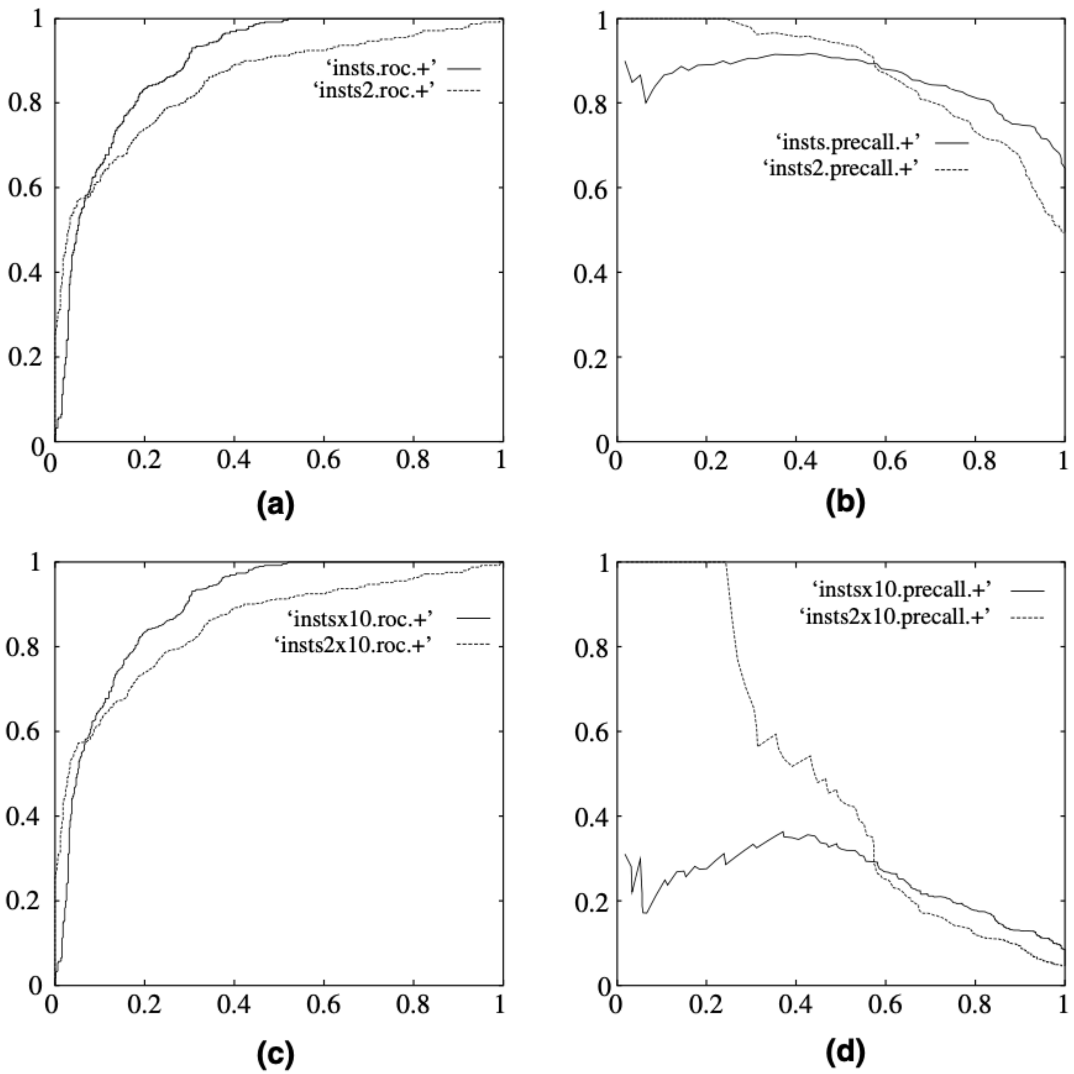



Fig. 5. ROC and precision-recall curves under class skew. (a) ROC curves, 1:1; (b) precision-recall curves, 1:1; (c) ROC curves, 1:10 and (d) precision-recall curves, 1:10.

# ROC vs PR Curves

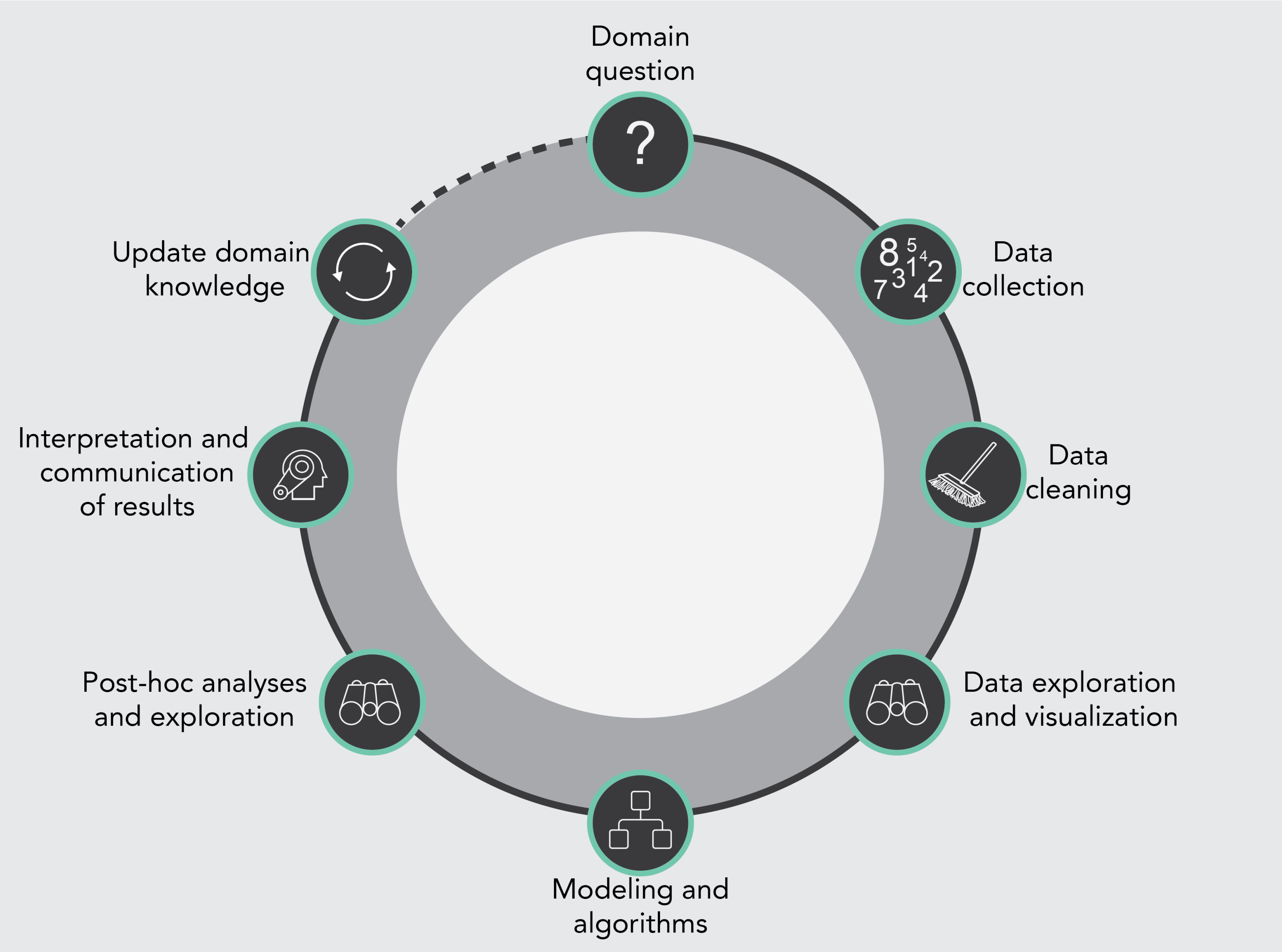
- Generally, precision-recall curves are preferred when there is class imbalance
- ROC curves tend to paint an overly optimistic view of the model on datasets with class imbalance
- PR calculations do not involve the true negatives rate and hence do not typically present such an optimistic view

# Other errors for regression

- Mean-squared error
- Absolute loss
- Correlation between true responses and predicted responses

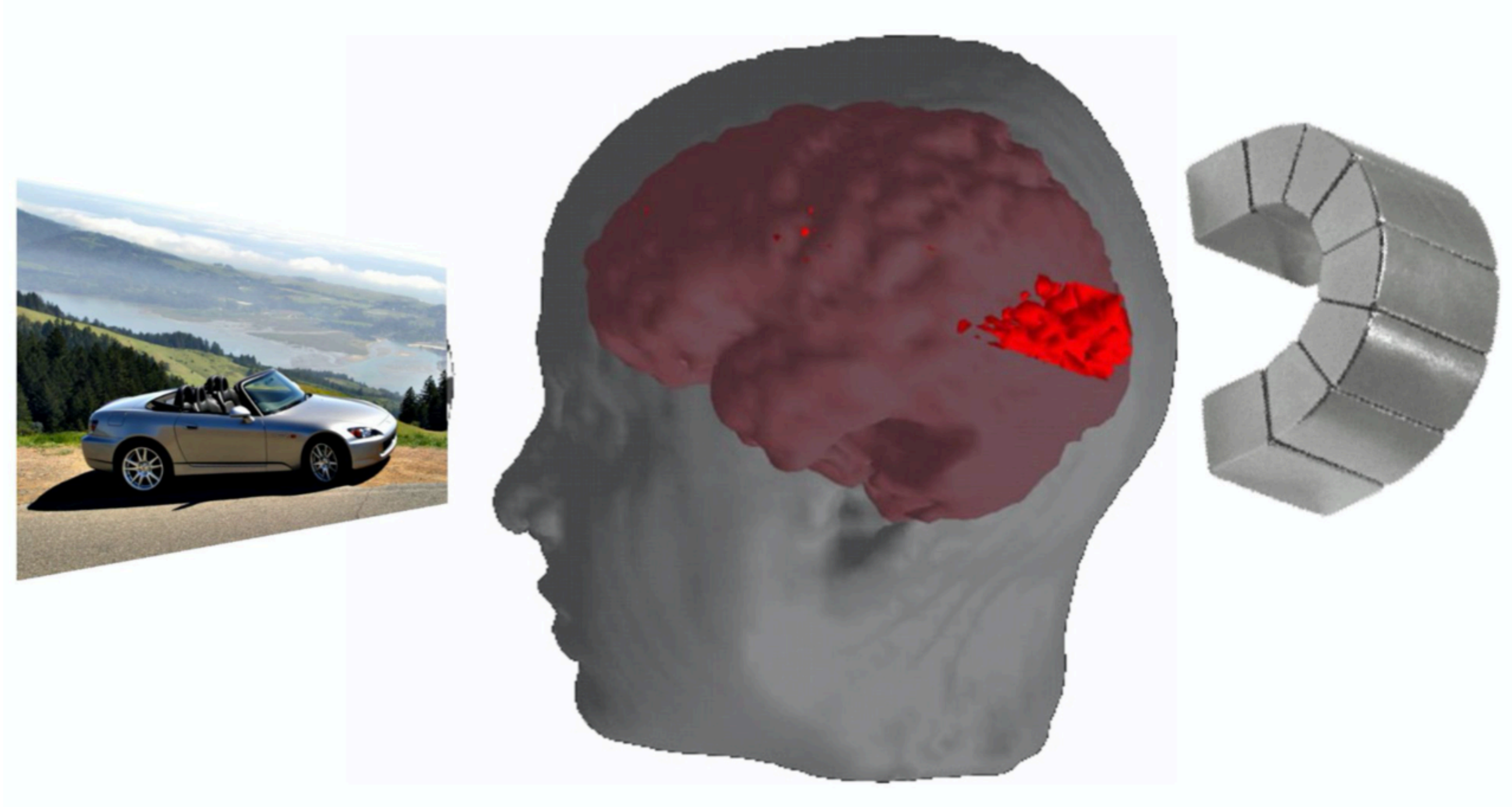
# STAT215A in a Nutshell

- Imagine you are a statistician working at PG&E and want to accurately predict when and where a fire will start
- In groups…
  - Walk through the data science life cycle and write down questions/things you would think about at each stage
  - How do core concepts from class/section fit into the data science life cycle?

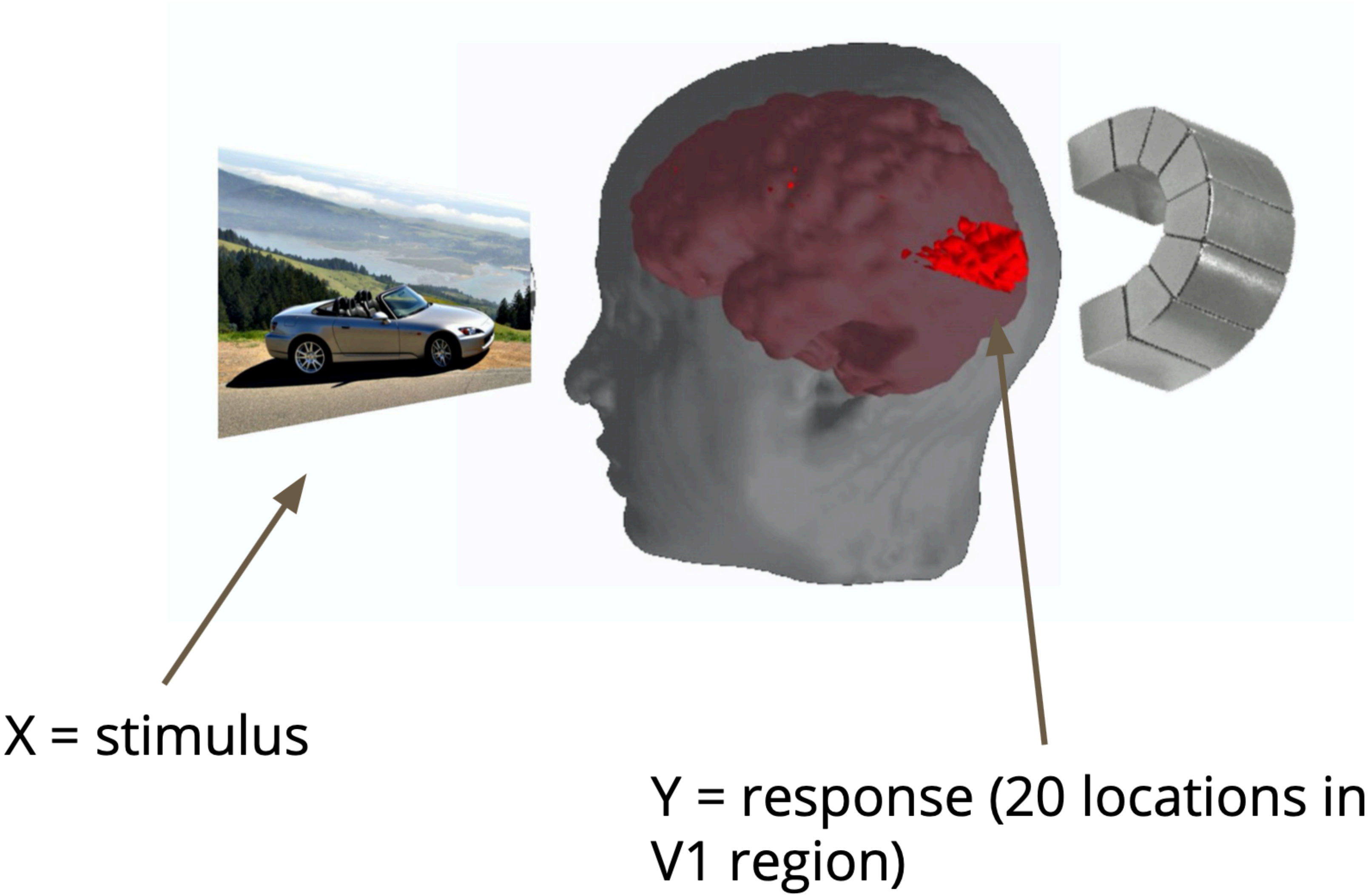


# Final Project (Due 11:59pm Friday, December 13)

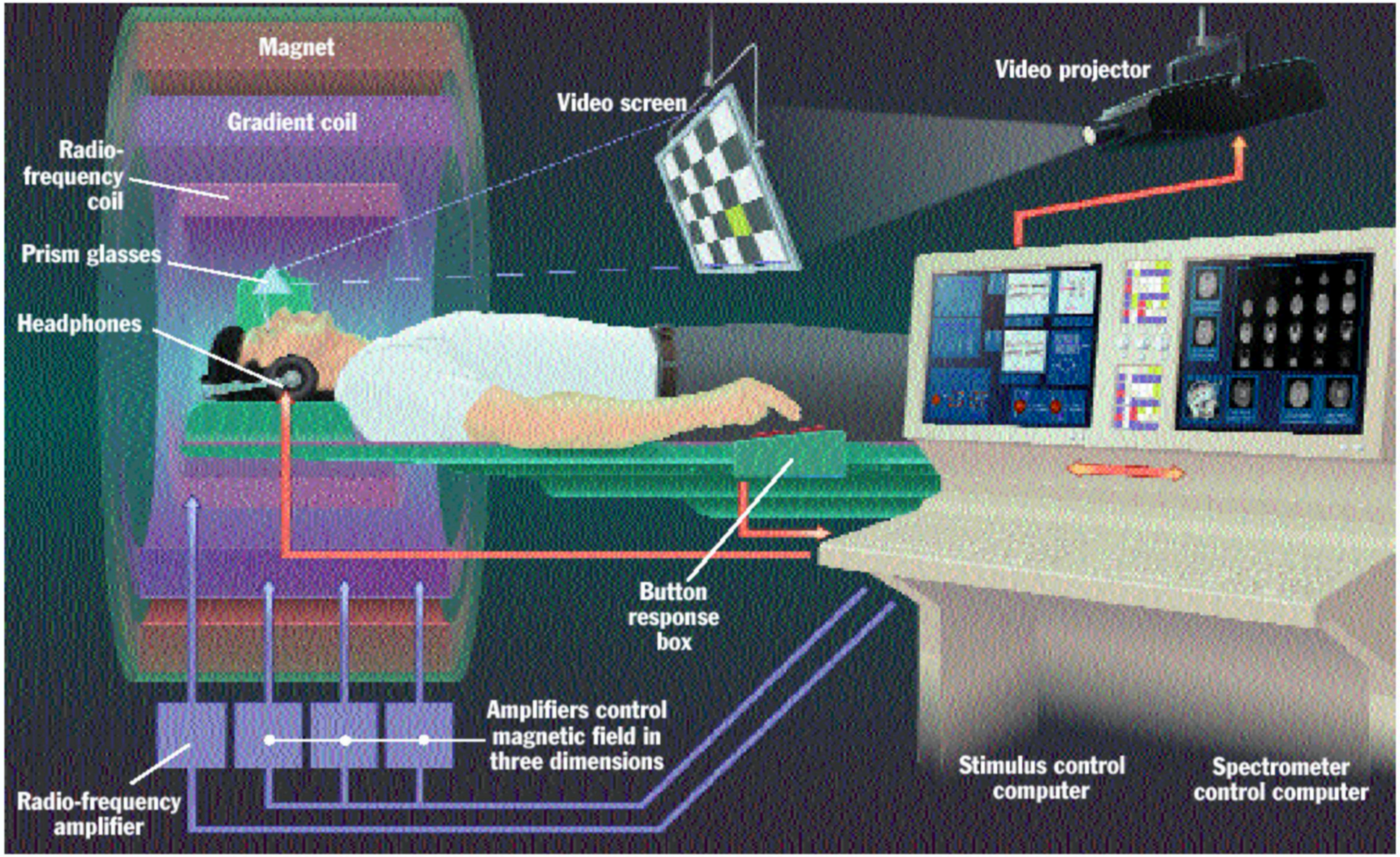
- **Goal:** predict the brain's response to images



# Final Project (Due 11:59pm Friday, December 13)

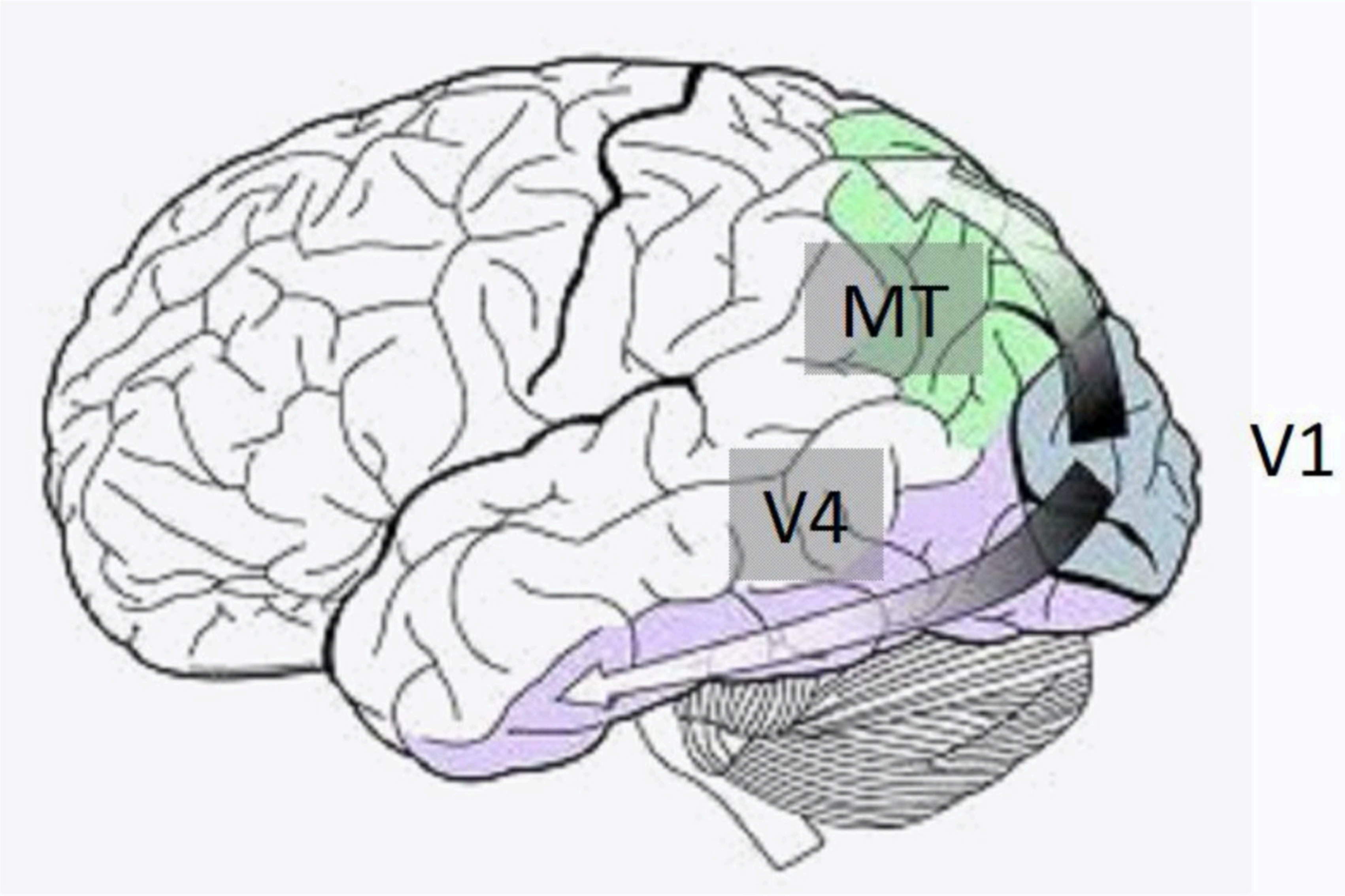


# fMRI

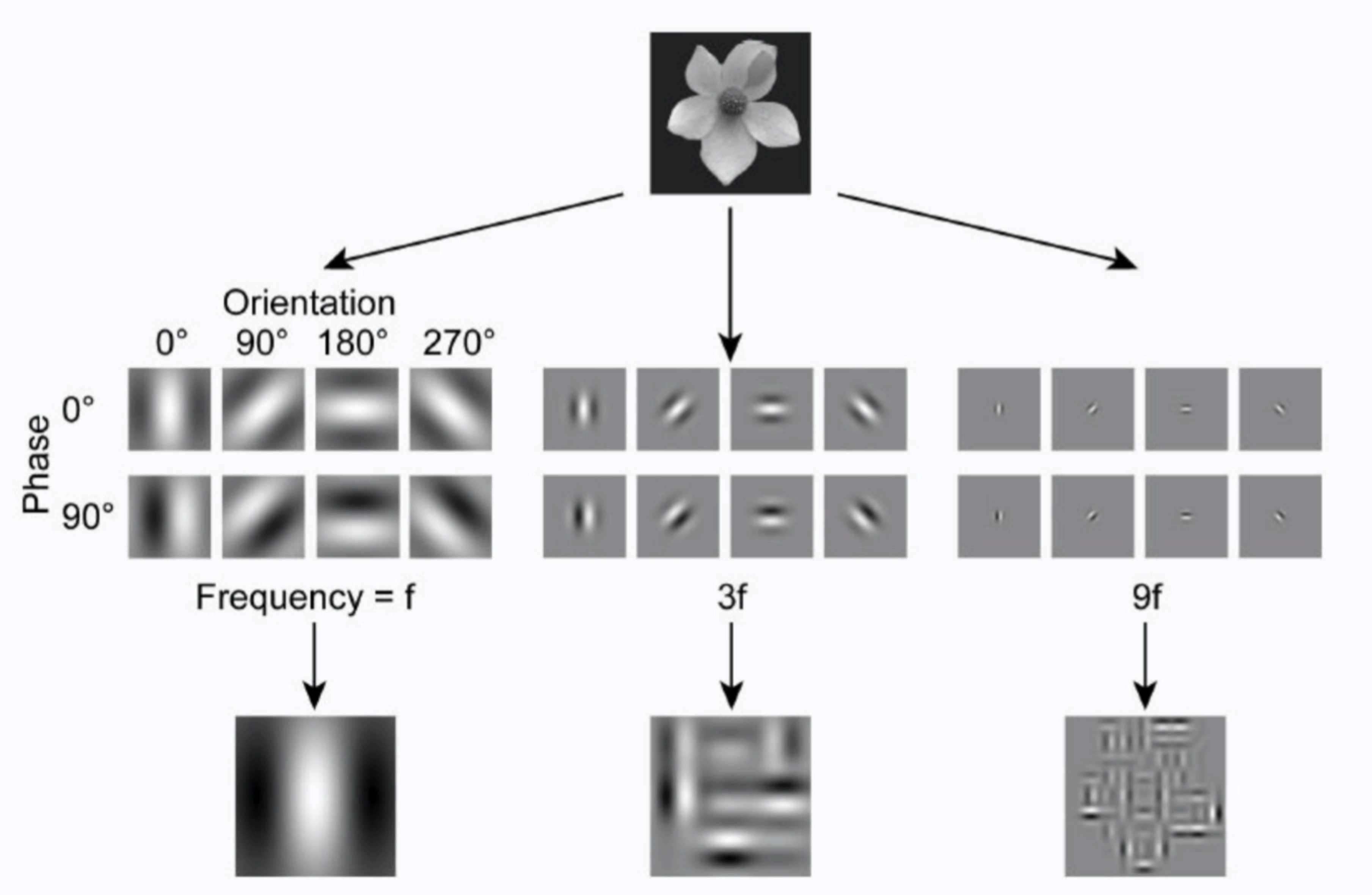


Magnet
Gradient coil
Radio-frequency coil
Prism glasses
Headphones
Video screen
Video projector
Button response box
Amplifiers control magnetic field in three dimensions
Radio-frequency amplifier
Stimulus control computer
Spectrometer control computer

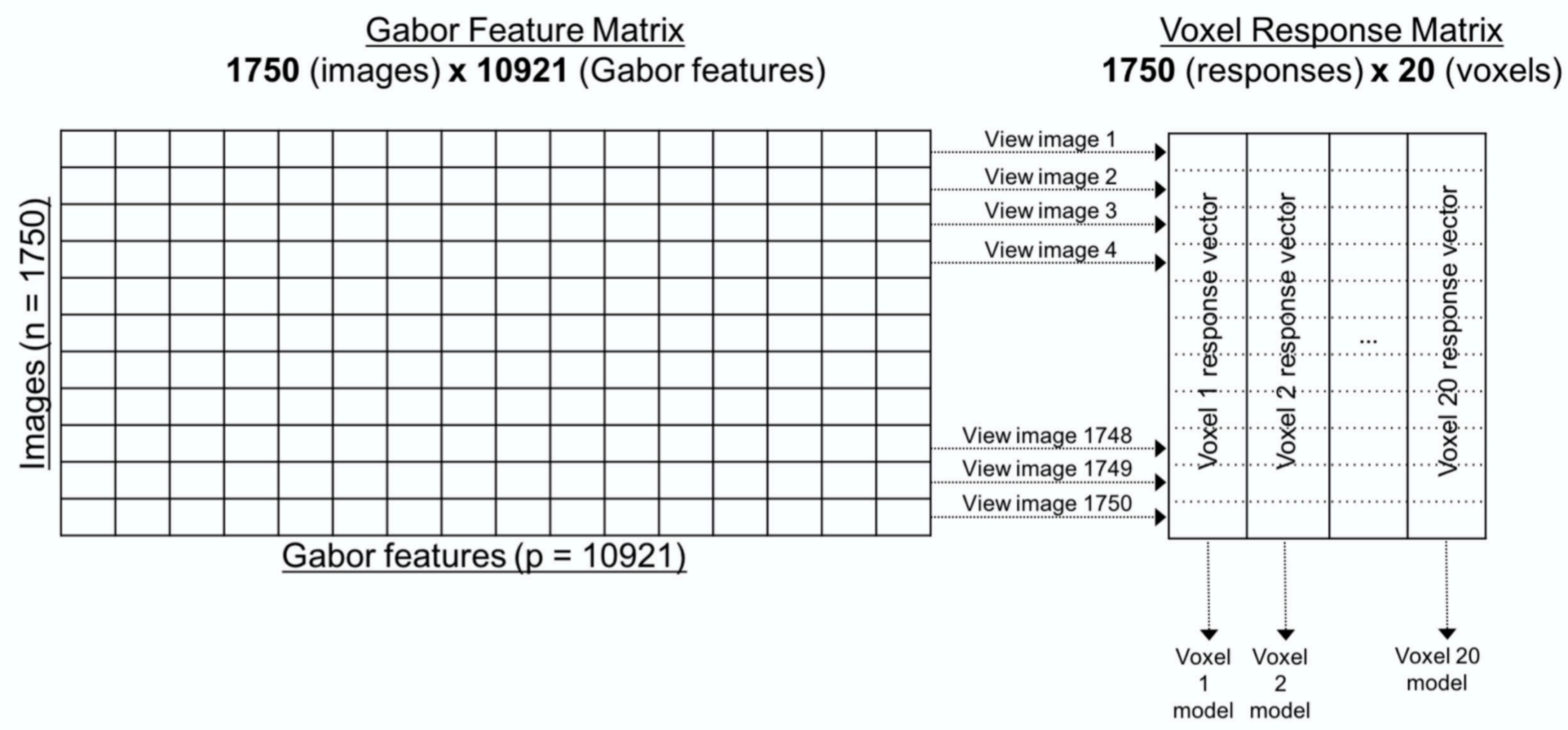
# V1 Region



# Gabor Transform



# The Data



Gabor Feature Matrix
1750 (images) x 10921 (Gabor features)
Voxel Response Matrix
1750 (responses) x 20 (voxels)
Images (n = 1750)
Gabor features (p = 10921)
View image 1
View image 2
View image 3
View image 4
View image 1748
View image 1749
View image 1750
Voxel 1 response vector
Voxel 2 response vector
...
Voxel 20 response vector
Voxel 1 model
Voxel 2 model
Voxel 20 model

# Your Tasks

- Guided analysis from domain formulation to post-hoc analysis
  - Predict the voxel responses to new images
  - Interpret the prediction models
- In addition to the guided questions given in the instructions, points will be allotted to
  - Visualizations
  - Introduction/discussion/conclusion
  - Overall formatting and presentation
    - Cross-referencing of figures and tables (see week2 lab_gapminder_solutions.Rmd), appropriate captions, no R output
  - Reproducibility
  - Readability of code

# Resources

- Looking at the data example: fmri_example.R
- Relevant papers:
  - Kay et al. Identifying natural images from human brain activity (Nature paper and supplemental materials)
  - Lim, Yu – ESCV paper
- Yuval Benjamini slide deck