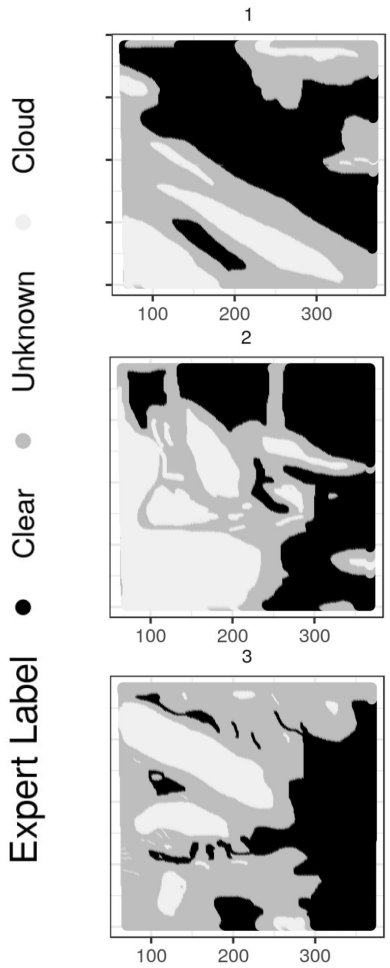
STAT 215A Fall 2020 Week 15

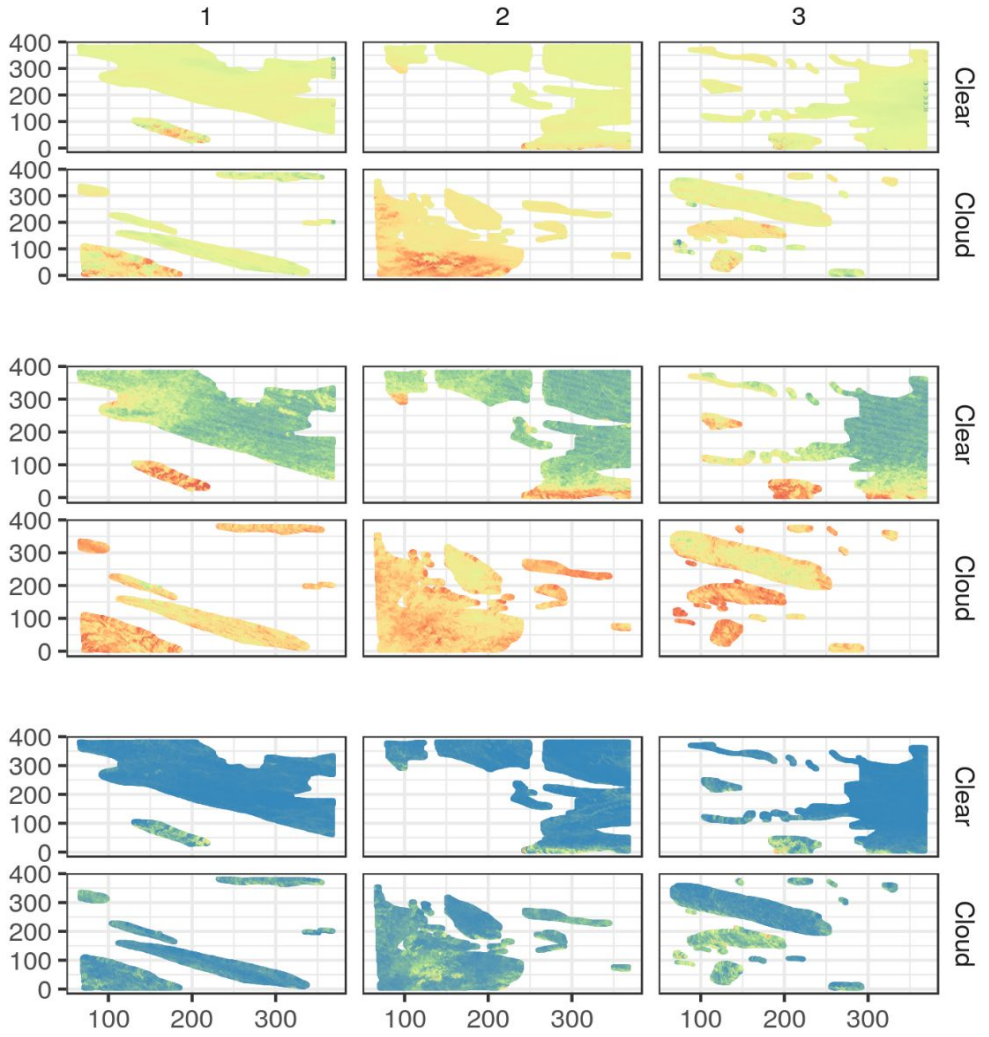
James Duncan, OH: M, Th 2-4pm

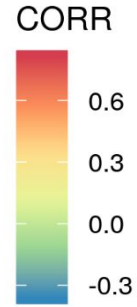
Announcements

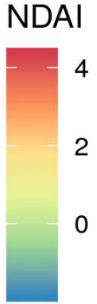
- Final project due in just over one week: **Friday, December 11 at 11:59pm**
- Lab 4 grades will be released later today (great work!)
- I will hold OH next week at the usual times and during the discussion section time.
- Bin wants to know if there's interest in informal final project presentations during finals week.
    - I would ask a few groups to talk about their report.

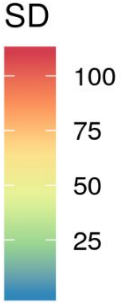
Lab 4 retrospective







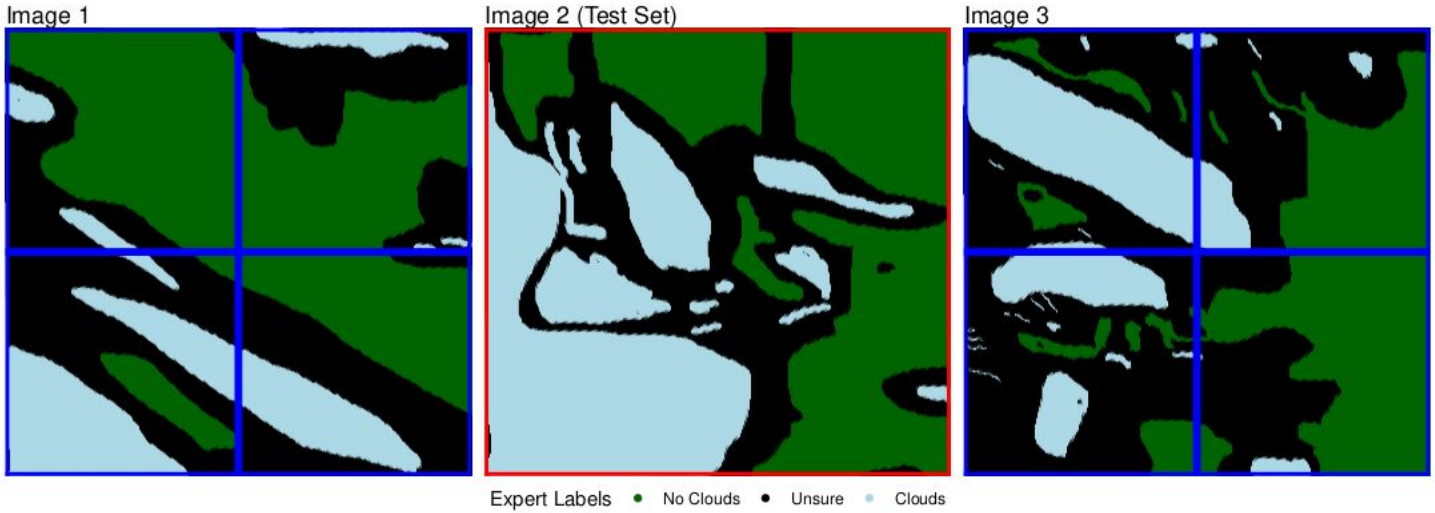




Lab 4 retrospective

RF performance with NDAI + CORR + SD for different sample splitting strategies:

- Completely random split?
- Blocked random split?
- Full image test set?



Lab 4 retrospective

Modeling setup: Random Forest with 80 trees, NDAI + CORR + SD

Unlabeled pixels?: Combine unlabeled and cloudy

Lab 4 retrospective

Modeling setup: Random Forest with 80 trees, NDAI + CORR + SD
Unlabeled pixels?: Combine unlabeled and cloudy

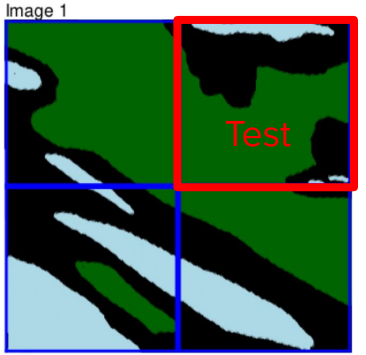
- Completely random split
  - 25% test
  - Accuracy: 0.874

Lab 4 retrospective

Modeling setup: Random Forest with 80 trees, NDAI + CORR + SD

Unlabeled pixels?: Combine unlabeled and cloudy

- Completely random split
  - 25% test
  - Accuracy: 0.874
- Blocked random split
  - Upper right quadrant of each image in test set
  - Accuracy: 0.851

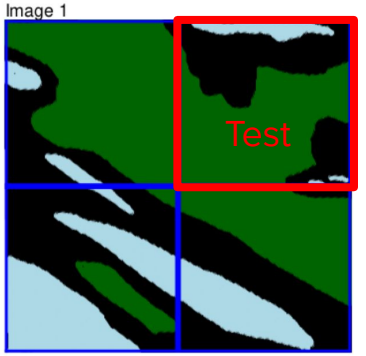


Lab 4 retrospective

Modeling setup: Random Forest with 80 trees, NDAI + CORR + SD

Unlabeled pixels?: Combine unlabeled and cloudy

- Completely random split
  - 25% test
  - Accuracy: 0.874
- Blocked random split
  - Upper right quadrant of each image in test set
  - Accuracy: 0.851
- Full image test set?
  - Image 1 accuracy: 0.812

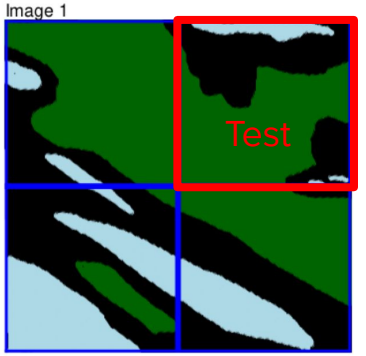


Lab 4 retrospective

Modeling setup: Random Forest with 80 trees, NDAI + CORR + SD

Unlabeled pixels?: Combine unlabeled and cloudy

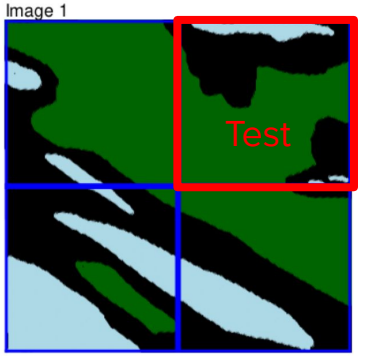
- Completely random split
  - 25% test
  - Accuracy: 0.874
- Blocked random split
  - Upper right quadrant of each image in test set
  - Accuracy: 0.851
- Full image test set?
  - Image 1 accuracy: 0.812
  - Image 2 accuracy: 0.888



Lab 4 retrospective

Modeling setup: Random Forest with 80 trees, NDAI + CORR + SD
Unlabeled pixels?: Combine unlabeled and cloudy

- Completely random split
  - 25% test
  - Accuracy: 0.874
- Blocked random split
  - Upper right quadrant of each image in test set
  - Accuracy: 0.851
- Full image test set?
  - Image 1 accuracy: 0.812
  - Image 2 accuracy: 0.888
  - Image 3 accuracy: 0.824



Lab 4 retrospective

Scenario: New data with no responses and you have to predict!

Try to use the data you have to mimic this scenario as closely as you can.

- This means we have to pretend the test set doesn't exist until our model is finalized.
- This means we have to think very carefully about sample splitting.
- Low error does not equal generalizability!

How are your final projects going?

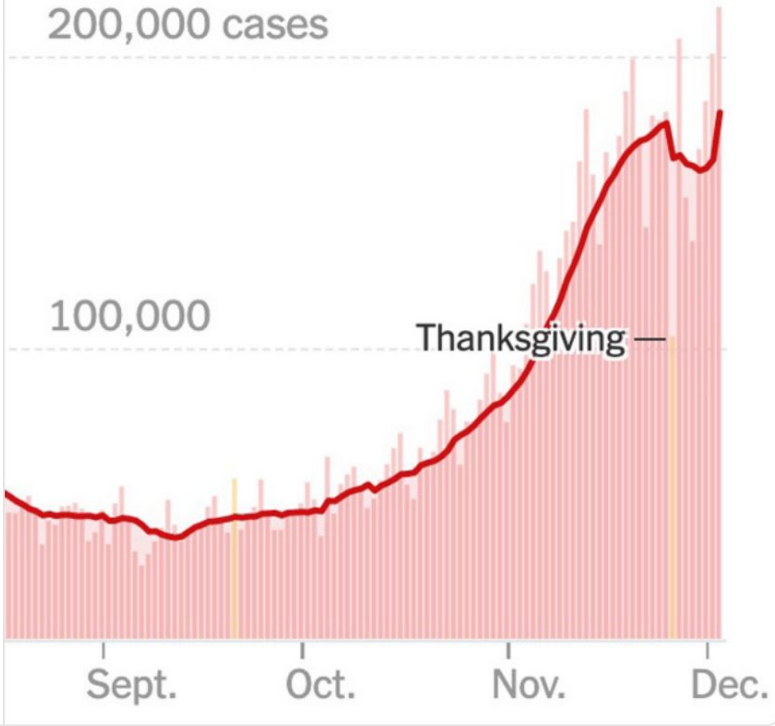
The State of the Virus This Week

Mitch Smith
Reporting on the coronavirus

It's been a terrible week for virus news. We began in a Thanksgiving data tunnel.

The national case curve had a deceptive glimmer of hope as many states continued to report suspiciously low numbers →

—
The New York Times



How are your final projects going?

The State of the Virus This Week

Mitch Smith
Reporting on the coronavirus

Our team of journalists tracking the virus is monitoring an uptick in **California**, where there were more than 21,000 cases on Thursday. That's the most any state has announced in a day.

Gov. Gavin Newsom said his state might run out of hospital beds by mid-December.

—
Bryan Denton for The New York Times

Tips for the final project

- Don’t forget the domain problem!
- Shoot for unbiased estimates of test error.
- You should have your end-to-end pipeline set up by now (or close).
  - This means ability to create predictions out to some horizon.
  - This doesn’t necessarily mean you should have every task withing the pipeline complete by now.
  - You can continually refine the pipeline, but don’t wait until the last minute to put everything together!
- Agree on a consistent visualization style with your groupmates.

Thank you!

Breakout rooms to meet with your final lab teammates + OH until 1pm