20 Oct 24

Add everything to github, literature, code etc

**Alicjia**

Classify species which ones do we want to focus on? In both training and datesets. 2 or 3 species build a classifier to what kind of species it is. 500 species in training data, choose 2-3 to reduce dimensions, answer question of relationships between species and locations.

**Gabin**

Ordinal embedding to reduce our dimensions (species) - apply to linear regressor, random forest decision trees for example

Ishita climates and coordinated

**Ishita**

Mapping historical data to areas vulnerable to climate change with future climate projections

Migratory patterns? Time stamps.

Alicjia cross validated data merging data sets for more reliable data – as a way of testing our method model is it more reliable than how the data was presented to us?

Gabin where the species is NOT as well as where it is –

Alicjia checked for null values – found none

Box plot any species to just show basic variation in distribution – show for species only that have large variablility

Should we scale/widen data standardise data?

Supervised/unsupervised models?

How to answer the question of relationships between species and locations?

1. Find region of interest e.g. 5 species of interest in same taxon? Build classifier theyr locations algorithm to rpredicts for new data point of location what species would be there? Biological significance of that
   1. Model performance comparison
   2. Translocation assessment
   3. Unsupervised learning identify few species of interest and use clustering to find regions where they are elbow plot for number of regions

We are big on clustering – what species exist with other species? What species is most likely to exist in a given cluster?

Pip

Look at lecture 1 (?) regions of where point was likeliest what the point would fall if it was in that region decision boundary plot!!

Not machine learning problem things can happen over time – each thing has probability

Coordinate

**27 Oct 24**

Pip – decision tree, very basic. Compare methods potentially? Definitely bring to progress report. How to visualise multiple species. Plot onto a map.

Gabin – random forest regressor, best fit decision trees. Takes in categorical features species ids one-hot encoding. Outputs single coordinate. Output as habitat? Ishita - Might have multiple wrong results? Map extra data with more context. Problem is see ocean bird picture – bird located in all parts of the world. Solve by sub-clusters?

Alicja – violin plots x = species, y = latitude/longitude

Clusters –

1. 1 species with sub population clusters
   1. Elbow plot – optimal number of clusters
   2. Plot map of hotspots

Add habitats for sub population clusters

1. All species data and find clusters
   1. Narrow down to Australia – 2 main clusters
      1. What species are clustered together?
      2. What species are most common in a given cluster?

Ishita – DBSCAN to cluster species with species

Overall narrative –

if a species can be in a predicted environment

Species in areas that are at risk of bein underwater

Based on current distribution of species

Classify species as at risk from climate change (for example in flood risk) 🡪 Make a dictionary from these species 🡪 show other areas in the world where this species would be viable to relocate (include habitat classifier, climate data)

Other clustering algorithms – used k-means but can use other

Show clusters that have similar ratios of species

Intro

Problem of allocating at risk species to new locations based on climate and habitat data.

Body

Classify species as at risk from climate change (for example in flood risk) 🡪 Make a dictionary from these species 🡪 show other areas in the world where this species would be viable to relocate (include habitat classifier, climate data, altitude)

Results

Map of viable location predictions, how far away from the equator

Discussion

Critic: Species interactions, only have positive observation data, how do we know if a species will cause damage in a new location?

**Progress report:**

Cleaning data, data exploration

What tools we’ve been using

What data we’ve been using

Successes, failures, how it will be useful to final report

Context – intro, conclusion, references

Questions – How much of our failures and our criticisms?, should we scale, standardise?

Figures