Project 1

Downscaling

Cristobal Spielmann (cs181)

2023-11-13

## Executive Summary

This report attempts to apply a downscaling model on both the precipitation data across the state of Texas and the ERA5 parameter of temperature data of air at 2m above the Earth’s surface. This will aim to increase the resolution of the precipitation data available by using the temperature parameter from 1980 to 2020 as a predictor for the precipitation predictand. The two models of this project are a GLM model and a KNN model.

## Exploratory Data Analysis

heatmap(  
precip\_lon,  
precip\_lat,  
precip[:, :, 15341]';  
xlabel="Longitude",  
ylabel="Latitude",  
title="Precipitation on $(precip\_time[15341])"  
)

heatmap(  
precip\_lon,  
precip\_lat,  
precip[:, :, 15341]';  
xlabel="Longitude",  
ylabel="Latitude",  
title="Precipitation on $(precip\_time[15341])"  
)

pca\_model = fit(PCA, temp\_mat\_train; maxoutdim=25, pratio=0.999);  
p1 = plot(  
principalvars(pca\_model) / var(pca\_model);  
xlabel="# of PCs",  
ylabel="Fraction of Variance Explained",  
label=false,  
title="Variance Explained"  
)  
p2 = plot(  
cumsum(principalvars(pca\_model)) / var(pca\_model);  
xlabel="# of PCs",  
ylabel="Fraction of Variance Explained",  
label=false,  
title="Cumulative Variance Explained Plot"  
)  
plot(p1, p2; layout=(1, 2), size=(800, 400))

Scatter plot

scatter(  
 temp\_mat\_train  
 precip\_train;  
 label=false,  
 xlabel="Temperature (deg K)",  
 ylabel="Precipitation",  
 legend=:topleft  
)

## Methods

I attempted to fit two models to this data. The first was a GLM as a simpler model and in assuming that neither the temperature nor the precipitation data would be normally distributed. The second model is a KNN analysis model based on the work done in Lab 6.

## Model Comparison

GLM Model

pca\_model = fit(PCA, temp\_mat\_train; maxoutdim=25, pratio=0.999);  
p1 = plot(  
principalvars(pca\_model) / var(pca\_model);  
xlabel="# of PCs",  
ylabel="Fraction of Variance Explained",  
label=false,  
title="Variance Explained"  
)  
p2 = plot(  
cumsum(principalvars(pca\_model)) / var(pca\_model);  
xlabel="# of PCs",  
ylabel="Fraction of Variance Explained",  
label=false,  
title="Cumulative Variance Explained Plot"  
)  
plot(p1, p2; layout=(1, 2), size=(800, 400))

scatter(  
 temp\_mat\_train  
 precip\_train;  
 label=false,  
 xlabel="Temperature (deg K)",  
 ylabel="Precipitation",  
 legend=:topleft  
)

KNN Model

function knn(X::AbstractMatrix, X\_i::AbstractVector, K::Int)::Tuple{Int,AbstractVector}  
dist = [euclidean\_distance(X\_i, X[j, :]) for j in 1:size(X, 1)]  
idx = nsmallest(dist, K)  
w = 1 ./ dist[idx]  
w ./= sum(w)

## Conclusion

This process was complicated from how the data was collected; the data and code in the Project quarto file shows how I attempted to coordinate multiple .nc files of temperature with the precipitation data. After performing a GLM model and KNN model of downscaling of both datasets, I was unable to to produce a plot or model for either.