CS4701_Regional_Climate_Forecast_with_ANN_final

May 23, 2021

1 CS4701: Regional Weather Forcast with ANN

by Yifan Zhao(yz348) and Kristina Tian (yt523)

1.1 Abstract

In this project, we are developing a weather forecasting tool that aims to predict key regional weather parameters based on historical data. Various Artificial Neural Network methods are used in comparison to extract trends and predict patterns of the regional weather sequences. Multiple data visualization and analysis techniques are employed to aid feature engineering, feature selection, and model selection processes. Validation and test methods are employed to evaluate and compare the performance of the tool.

1.1.1 External Libraries

```
In [485]: import numpy as np
                                              import pandas as pd
                                              import matplotlib.pyplot as plt
                                             from netCDF4 import Dataset
                                              import xarray as xr
                                              # for ML
                                             from sklearn.preprocessing import StandardScaler, MinMaxScaler, RobustScaler, Polynomial Control of the Standard Scaler, Polynomial Control of 
                                             from keras.models import Sequential
                                             from keras.layers import LSTM, Dense , Dropout, Bidirectional
                                             from sklearn.metrics import mean_squared_error
                                              ## for plotting
                                             import matplotlib.pyplot as plt
                                             import seaborn as sns
                                              import statsmodels.api as sm
                                             from statsmodels.tsa.stattools import adfuller
                                             import warnings
                                             warnings.filterwarnings("ignore")
```

1.2 1. Data

1.2.1 1.1 Basic data

This is a data gateway for downloading customized GCMs (Global Climate Models)datasets, downsacled using MACA method (Multivariate Adaptive Constructed Analogs). This method enables us to retrive historical metrological datasets as far as 1950 and with a resolution up to 4km. *This gateway is currently not in use by this project, where we instead directly import MACA datasets about the Finger Lakes AVA.

```
In [ ]: lat_target=45.0
        lon_target=360-117.0
In []: pathname = 'http://thredds.northwestknowledge.net:8080/thredds/dodsC/agg_macav2metdata
In [ ]: filehandle=Dataset(pathname, 'r', format="NETCDF4")
        lathandle=filehandle.variables['lat']
        lonhandle=filehandle.variables['lon']
        timehandle=filehandle.variables['time']
        datahandle=filehandle.variables['specific_humidity']
In [ ]: time_num=365
        timeindex=range(0,time_num,1)
        time=timehandle[timeindex]
        lat = lathandle[:]
        lon = lonhandle[:]
In [ ]: #find indices of target lat/lon/day
        lat_index = (np.abs(lat-lat_target)).argmin()
        lon_index = (np.abs(lon-lon_target)).argmin()
        #check final is in right bounds
        if(lat[lat_index]>lat_target):
                if(lat_index!=0):
                        lat_index = lat_index - 1
        if(lat[lat_index]<lat_target):</pre>
                if(lat_index!=len(lat)):
                        lat_index =lat_index +1
        if(lon[lon_index]>lon_target):
                if(lon_index!=0):
                        lon_index = lon_index - 1
        if(lon[lon_index]<lon_target):</pre>
                if(lon_index!=len(lon)):
                        lon_index = lon_index + 1
        lat=lat[lat_index]
        lon=lon[lon_index]
In []: data = datahandle[timeindex,lat_index,lon_index]
In [ ]: days = np.arange(0,len(time))
        fig = plt.figure()
```

```
ax = fig.add_subplot(111)
ax.set_xlabel(u'Day of Year')
ax.set_ylabel(u'Specific Humidity(kg/kg)')
ax.set_title(u'1950 Daily Specific Humidity(BNU-ESM) ,\n %4.2f\u00b0N, %4.2f\u00b0W' %
ax.ticklabel_format(style='plain')
ax.plot(days,data,'b-')
plt.savefig("myPythonGraph.png")
plt.show()
```

1.2.2 1.2 Import Data

Import MACA datasets about the Finger Lakes AVA regional GCMs.

```
In [3]: # Fucntion for reading datasets
        def read_file(file):
            df = pd.read_csv(file, sep=',',header=0, encoding='unicode_escape')
            return df
        # Encode date to the number of days from 1/1/2020
        def encode_date(df):
            return (pd.to_datetime(df['date']).rsub(pd.Timestamp('2006/1/1')).dt.days)*(-1)
In [489]: from google.colab import drive
          import os
          drive.mount('/content/drive', force_remount=True)
          path = os.path.join(os.getcwd(), "drive", "My Drive", "Regional-Climate-Forecast-with
In []: df_his_daily = read_file(os.path.join(path, 'MACA_41.9795 Latitude, -76.9813 Longitude)
       df_cur_daily = read_file(os.path.join(path,'MACA_41.9795 Latitude, -76.9813 Longitude_
        df_future_daily = read_file(os.path.join(path, 'MACA_41.9795 Latitude, -76.9813 Longitue
        df_his_daily['date'] = encode_date(df_his_daily)
In [478]: #Without Google Colab
          df_his_daily = read_file('MACA_41.9795 Latitude, -76.9813 Longitude_2006-2019.csv')
          df_cur_daily = read_file('MACA_41.9795 Latitude, -76.9813 Longitude_2019-2020.csv')
          df_future_daily = read_file('MACA_41.9795 Latitude, -76.9813 Longitude_2020-2021.csv
          # df_his_daily['date'] = encode_date(df_his_daily)
```

1.3 2. Feature Engineering

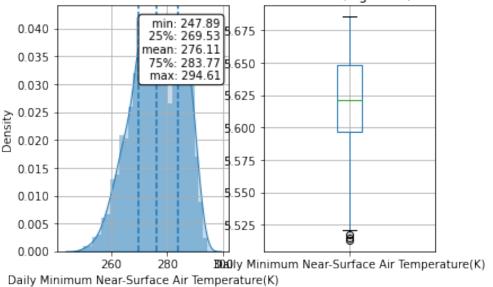
In this section, we analyze and visualize different features of the data to get a sense of what we are working with. These visualization show the distinct patterns of different features in the training set from 2006 to 2019. It also shows where anomalies are present, which will be addressed during normalization.

```
cur_pcp = df_cur_daily[['Precipitation(mm)']]
          cur_max_temp = df_cur_daily[['Daily Maximum Near-Surface Air Temperature(K)']]
          future_pcp = df_future_daily[['Precipitation(mm)']]
          future_max_temp = df_future_daily[['Daily Maximum Near-Surface Air Temperature(K)']]
          # Cut off the last year
          his_pcp = his_pcp[4748:]
          his_max_temp = his_max_temp[4748:]
          cur_pcp = cur_pcp[:365]
          cur_max_temp = cur_max_temp[:365]
          future_pcp = future_pcp[:365]
          future_max_temp = future_max_temp[:365]
In [311]: # functions for visualizing numerical data distributions
          def check_distribution_conti(dtf, x):
              fig, ax = plt.subplots(nrows=1, ncols=2, sharex=False, sharey=False)
              fig.suptitle(x, fontsize=20)
              ### distribution
              ax[0].title.set_text('distribution')
              variable = dtf[x].fillna(dtf[x].mean())
              breaks = np.quantile(variable, q=np.linspace(0, 1, 11))
              variable = variable[ (variable > breaks[0]) & (variable <</pre>
                                  breaks[10]) ]
              sns.distplot(variable, hist=True, kde=True, kde_kws={"shade": True}, ax=ax[0])
              des = dtf[x].describe()
              ax[0].axvline(des["25%"], ls='--')
              ax[0].axvline(des["mean"], ls='--')
              ax[0].axvline(des["75%"], ls='--')
              ax[0].grid(True)
              des = round(des, 2).apply(lambda x: str(x))
              box = '\n'.join(("min: "+des["min"], "25%: "+des["25%"], "mean: "+des["mean"], "
              ax[0].text(0.95, 0.95, box, transform=ax[0].transAxes, fontsize=10, va='top', has
              ### boxplot
              ax[1].title.set_text('outliers (log scale)')
              tmp_dtf = pd.DataFrame(dtf[x])
              tmp_dtf[x] = np.log(tmp_dtf[x])
              tmp_dtf.boxplot(column=x, ax=ax[1])
              plt.show()
          # functions for visualizing bivariant distribution between a categorical feature and
          def check_bivariant_cat(dtf, cat, num):
              fig, ax = plt.subplots(nrows=1, ncols=3, sharex=False, sharey=False)
              fig.suptitle(cat+" vs "+num, fontsize=20)
              ### distribution
              ax[0].title.set_text('density')
              for i in dtf[cat].unique():
                  sns.distplot(dtf[dtf[cat]==i][num], hist=False, label=i, ax=ax[0])
```

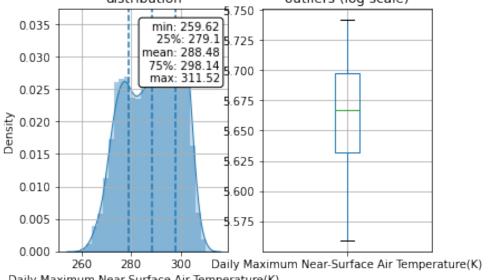
```
ax[0].grid(True)
              ### stacked
              ax[1].title.set_text('bins')
              breaks = np.quantile(dtf[num], q=np.linspace(0,1,11))
              tmp = dtf.groupby([cat, pd.cut(dtf[num], breaks, duplicates='drop')]).size().uns
              tmp = tmp[dtf[cat].unique()]
              tmp["tot"] = tmp.sum(axis=1)
              for col in tmp.drop("tot", axis=1).columns:
                   tmp[col] = tmp[col] / tmp["tot"]
              tmp.drop("tot", axis=1).plot(kind='bar', stacked=True, ax=ax[1], legend=False, g
              ### boxplot
              ax[2].title.set_text('outliers')
              sns.catplot(x=cat, y=num, data=dtf, kind="box", ax=ax[2])
              ax[2].grid(True)
              plt.show()
In [471]: titles = [
              "Daily Minimum Near-Surface Air Temperature(K)",
              "Daily Maximum Near-Surface Air Temperature(K)",
              "Daily Mean Near-Surface Specific Humidity(kgkg-1)",
              "vpd(kPa)",
              "Precipitation(mm)",
          1
          feature_keys = [
              "Daily Minimum Near-Surface Air Temperature(K)",
              "Daily Maximum Near-Surface Air Temperature(K)",
              "Daily Mean Near-Surface Specific Humidity(kgkg-1)",
              "vpd(kPa)",
              "Precipitation(mm)",
          ]
          colors = [
              "blue".
              "orange",
              "green",
              "red",
              "purple",
          ]
          # Function for visualizing data distribution
          def show_raw_visualization(data):
              time_data = data["date"]
              fig, axes = plt.subplots(
                  nrows=3, ncols=2, figsize=(15, 20), dpi=80, facecolor="w", edgecolor="k"
              )
              for i in range(len(feature_keys)):
```

```
key = feature_keys[i]
                  c = colors[i % (len(colors))]
                  t_data = data[key]
                  t_data.index = time_data
                  t_data.head()
                  ax = t_data.plot(
                      ax=axes[i // 2, i % 2],
                      color=c,
                      title="{} - {}".format(titles[i], key),
                      rot=25,
                  )
                  ax.legend([titles[i]])
              plt.tight_layout()
          # Function for visualizing correlation heatmap
          def show_heatmap(data):
              plt.matshow(data.corr())
              plt.xticks(range(data.shape[1]), data.columns, fontsize=14, rotation=90)
              plt.gca().xaxis.tick_bottom()
              plt.yticks(range(data.shape[1]), data.columns, fontsize=14)
              cb = plt.colorbar()
              cb.ax.tick_params(labelsize=14)
              plt.title("Feature Correlation Heatmap", fontsize=14)
              plt.show()
In [481]: check_distribution_conti(df_his_daily, 'Daily Minimum Near-Surface Air Temperature(K
          check_distribution_conti(df_his_daily, 'Daily Maximum Near-Surface Air Temperature(K
          check_distribution_conti(df_his_daily, 'Daily Mean Near-Surface Specific Humidity(kg
          check_distribution_conti(df_his_daily, 'vpd(kPa)')
          check_distribution_conti(df_his_daily, 'Precipitation(mm)')
```

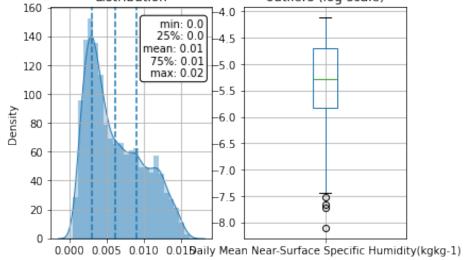
Daily Minimum Near-Surface Air Temperature (K)



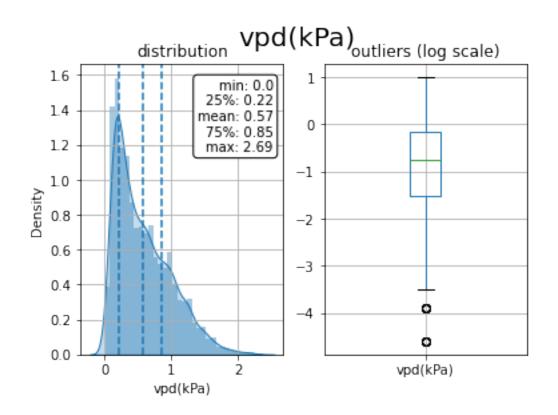
Daily Maximum Near-Surface Air Temperature (K)

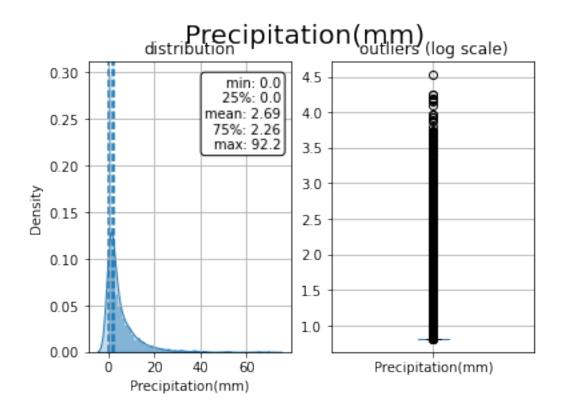


Daily Mean Near-Surface Specific Humidity(kgkg-1)

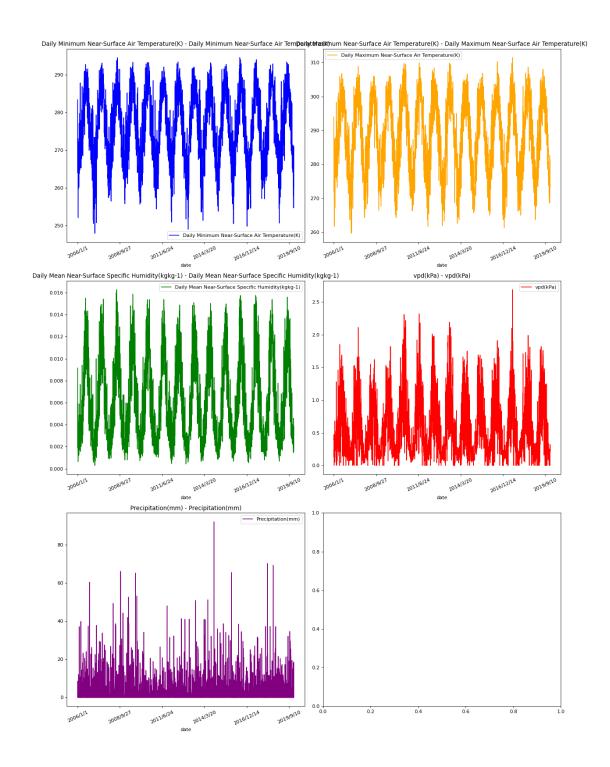


Daily Mean Near-Surface Specific Humidity(kgkg-1)

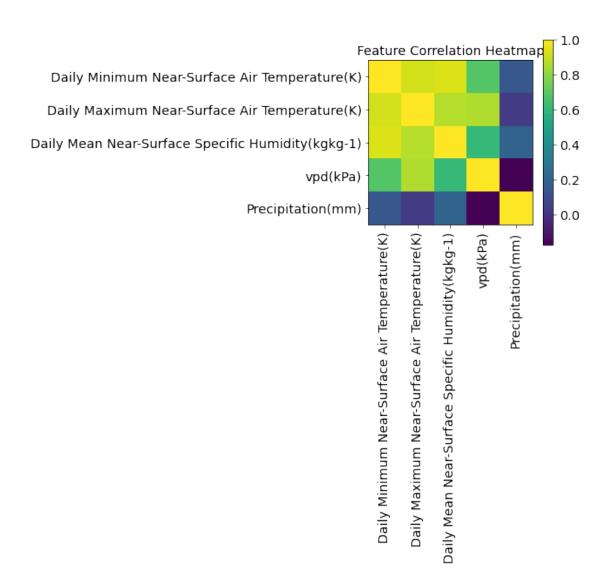


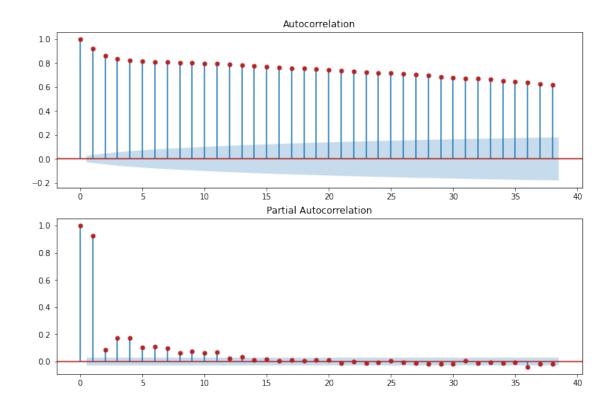


In [469]: show_raw_visualization(df_his_daily)



In [480]: show_heatmap(df_his_daily.drop('date', 1))





1.4 3.Train RNN for Predicting Tempreture

In this section, we train and evaluate a Pytorch RNN model in predicting Daily Maximum Near-Surface Air Temperature(K)

```
In [332]: import torch
    import torch.nn as nn
```

1.4.1 3.1 Data preprocessing

In [443]: # Data normalization

```
sc = MinMaxScaler(feature_range=(0,1))
In [444]: train_dataset = his_max_temp
          testdataset = cur_max_temp
In [445]: train_dataset = sc.fit_transform(train_dataset)
          x_train = []
          y_train = []
          n_future = 1
          n_past = 30 # Past 30 days
          for i in range(0,len(train_dataset)-n_past-n_future+1):
              x_train.append(train_dataset[i : i + n_past ])
              y_train.append(train_dataset[i + n_past : i + n_past + n_future ])
          x_train , y_train = np.array(x_train), np.array(y_train)
In [446]: # Preprocessing test data
          testdataset = sc.transform(testdataset)
          x_test = []
          y_test = []
          n_future = 1
          n past = 30
          for i in range(0,len(testdataset)-n_past-n_future+1):
              x_test.append(testdataset[i : i + n_past])
              y_test.append(testdataset[i + n_past : i + n_past + n_future, 0])
          x_test , y_test = np.array(x_test), np.array(y_test)
          x_test = np.reshape(x_test, (x_test.shape[0] , x_test.shape[1], 1) )
In [447]: x_test_tensor = torch.Tensor(x_test)
          y_test_tensor = torch.Tensor(y_test)
          x_train_tensor = torch.Tensor(x_train)
          y_train_tensor = torch.Tensor(y_train)
In [429]: x_train_tensor.shape
Out[429]: torch.Size([335, 30, 1])
1.4.2 3.2 Train model
In [430]: class RNN(nn.Module):
                  def __init__(self, input_dim, hidden_dim, layer_dim, output_dim): # Add rele
                          super(RNN, self).__init__()
                          self.input_dim = input_dim
                          self.hidden_dim = hidden_dim
                          self.layer_dim = layer_dim
                          self.output_dim = output_dim
                          self.rnn = nn.RNN(input_dim, hidden_dim, layer_dim, batch_first=True
```

```
self.lastlayer = nn.Linear(hidden_dim, output_dim)
                          self.softmax = nn.LogSoftmax(dim=1)
                          self.loss = nn.MSELoss()
                  def compute Loss(self, predicted vector, gold label):
                          return self.loss(predicted_vector.float(), gold_label.float())
                  def forward(self, inputs):
                          output, final_hidden = self.rnn(inputs)
                          final_output = self.lastlayer(output).squeeze(-1)
                          predicted_vector = torch.mode(final_output,1).values
                          return predicted_vector
                  def load_model(self, save_path):
                          self.load_state_dict(torch.load(save_path))
                  def save_model(self, save_path):
                          torch.save(self.state_dict(), save_path)
In [431]: model = RNN(input_dim=1, output_dim=1, hidden_dim=12, layer_dim=1)
          n = pochs = 1000
          1r=0.1
          criterion = nn.MSELoss()
          optimizer = torch.optim.Adam(model.parameters(), lr=lr)
          for epoch in range(1, n epochs + 1):
              optimizer.zero_grad()
              output = model(x_train_tensor)
              y_train_tensor = y_train_tensor.view(-1)
              loss = criterion(output.float(), y_train_tensor.float())
              loss.backward()
              optimizer.step()
              if epoch\%10 == 0:
                  print('Epoch: {}/{}.....'.format(epoch, n_epochs), end=' ')
                  print("Loss: {:.4f}".format(loss.item()))
Epoch: 10/1000... Loss: 0.1007
Epoch: 20/1000... Loss: 0.0585
Epoch: 30/1000... Loss: 0.0398
Epoch: 40/1000... Loss: 0.0337
Epoch: 50/1000... Loss: 0.0298
Epoch: 60/1000... Loss: 0.0281
Epoch: 70/1000... Loss: 0.0279
Epoch: 80/1000... Loss: 0.0277
Epoch: 90/1000... Loss: 0.0276
Epoch: 100/1000... Loss: 0.0276
Epoch: 110/1000... Loss: 0.0276
Epoch: 120/1000... Loss: 0.0276
```

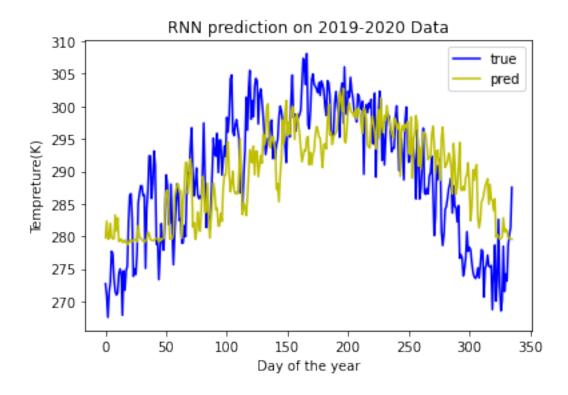
```
Epoch: 130/1000... Loss: 0.0276
Epoch: 140/1000... Loss: 0.0276
Epoch: 150/1000... Loss: 0.0276
Epoch: 160/1000... Loss: 0.0276
Epoch: 170/1000... Loss: 0.0276
Epoch: 180/1000... Loss: 0.0276
Epoch: 190/1000... Loss: 0.0276
Epoch: 200/1000... Loss: 0.0276
Epoch: 210/1000... Loss: 0.0276
Epoch: 220/1000... Loss: 0.0276
Epoch: 230/1000... Loss: 0.0276
Epoch: 240/1000... Loss: 0.0276
Epoch: 250/1000... Loss: 0.0276
Epoch: 260/1000... Loss: 0.0276
Epoch: 270/1000... Loss: 0.0276
Epoch: 280/1000... Loss: 0.0276
Epoch: 290/1000... Loss: 0.0276
Epoch: 300/1000... Loss: 0.0276
Epoch: 310/1000... Loss: 0.0276
Epoch: 320/1000... Loss: 0.0276
Epoch: 330/1000... Loss: 0.0276
Epoch: 340/1000... Loss: 0.0276
Epoch: 350/1000... Loss: 0.0276
Epoch: 360/1000... Loss: 0.0276
Epoch: 370/1000... Loss: 0.0276
Epoch: 380/1000... Loss: 0.0276
Epoch: 390/1000... Loss: 0.0276
Epoch: 400/1000... Loss: 0.0276
Epoch: 410/1000... Loss: 0.0276
Epoch: 420/1000... Loss: 0.0276
Epoch: 430/1000... Loss: 0.0276
Epoch: 440/1000... Loss: 0.0276
Epoch: 450/1000... Loss: 0.0276
Epoch: 460/1000... Loss: 0.0276
Epoch: 470/1000... Loss: 0.0276
Epoch: 480/1000... Loss: 0.0276
Epoch: 490/1000... Loss: 0.0276
Epoch: 500/1000... Loss: 0.0276
Epoch: 510/1000... Loss: 0.0276
Epoch: 520/1000... Loss: 0.0276
Epoch: 530/1000... Loss: 0.0276
Epoch: 540/1000... Loss: 0.0276
Epoch: 550/1000... Loss: 0.0276
Epoch: 560/1000... Loss: 0.0276
Epoch: 570/1000... Loss: 0.0276
Epoch: 580/1000... Loss: 0.0276
Epoch: 590/1000... Loss: 0.0276
Epoch: 600/1000... Loss: 0.0276
```

```
Epoch: 610/1000... Loss: 0.0276
Epoch: 620/1000... Loss: 0.0276
Epoch: 630/1000... Loss: 0.0276
Epoch: 640/1000... Loss: 0.0276
Epoch: 650/1000... Loss: 0.0276
Epoch: 660/1000... Loss: 0.0276
Epoch: 670/1000... Loss: 0.0276
Epoch: 680/1000... Loss: 0.0276
Epoch: 690/1000... Loss: 0.0276
Epoch: 700/1000... Loss: 0.0276
Epoch: 710/1000... Loss: 0.0276
Epoch: 720/1000... Loss: 0.0276
Epoch: 730/1000... Loss: 0.0276
Epoch: 740/1000... Loss: 0.0276
Epoch: 750/1000... Loss: 0.0276
Epoch: 760/1000... Loss: 0.0276
Epoch: 770/1000... Loss: 0.0276
Epoch: 780/1000... Loss: 0.0276
Epoch: 790/1000... Loss: 0.0276
Epoch: 800/1000... Loss: 0.0276
Epoch: 810/1000... Loss: 0.0276
Epoch: 820/1000... Loss: 0.0276
Epoch: 830/1000... Loss: 0.0276
Epoch: 840/1000... Loss: 0.0276
Epoch: 850/1000... Loss: 0.0276
Epoch: 860/1000... Loss: 0.0276
Epoch: 870/1000... Loss: 0.0276
Epoch: 880/1000... Loss: 0.0276
Epoch: 890/1000... Loss: 0.0276
Epoch: 900/1000... Loss: 0.0276
Epoch: 910/1000... Loss: 0.0276
Epoch: 920/1000... Loss: 0.0276
Epoch: 930/1000... Loss: 0.0276
Epoch: 940/1000... Loss: 0.0276
Epoch: 950/1000... Loss: 0.0276
Epoch: 960/1000... Loss: 0.0276
Epoch: 970/1000... Loss: 0.0276
Epoch: 980/1000... Loss: 0.0276
Epoch: 990/1000... Loss: 0.0276
Epoch: 1000/1000... Loss: 0.0276
```

1.4.3 3.3 Test and Evaluation

```
mse = mean_squared_error(y_test, pred)
print('MSE for Tempreture Prediction using RNN: ', mse)
plt.plot(y_test, color = 'b', label = "true")
plt.plot(pred, color = 'y', label = "pred")
plt.ylabel('Tempreture(K)')
plt.xlabel('Day of the year')
plt.title("RNN prediction on 2019-2020 Data")
plt.legend()
plt.show()
```

MSE for Tempreture Prediction using RNN: 54.48186965591475



1.5 4. Train LSTM for Predicting Tempreture

In this section, we train and evaluate a LSTM RNN model in predicting Daily Maximum Near-Surface Air Temperature(K)

1.5.1 4.1 Data preprocessing

```
n_past = 30 # 30 past days
     training_set = his_max_temp
     training_set = sc.fit_transform(training_set)
     for i in range(0,len(training_set)-n_past-n_future+1):
        x_train.append(training_set[i : i + n_past, 0])
        y_train.append(training_set[i + n_past : i + n_past + n_future, 0])
     x_train , y_train = np.array(x_train), np.array(y_train)
     x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))
1.5.2 4.2 Train Model
In [236]: # Train LSTM model
     regressor = Sequential()
     regressor.add(Bidirectional(LSTM(units=30, return_sequences=True, input_shape = (x_t
     regressor.add(Dropout(0.2))
     regressor.add(LSTM(units= 30 , return_sequences=True))
     regressor.add(Dropout(0.2))
     regressor.add(LSTM(units= 30 , return_sequences=True))
     regressor.add(Dropout(0.2))
     regressor.add(LSTM(units= 30))
     regressor.add(Dropout(0.2))
     regressor.add(Dense(units = n_future,activation='linear'))
     regressor.compile(optimizer='adam', loss='mean_squared_error',metrics=['acc'])
     regressor.fit(x_train, y_train, epochs=500,batch_size=32 )
Epoch 1/500
Epoch 2/500
Epoch 3/500
Epoch 5/500
Epoch 6/500
Epoch 7/500
Epoch 8/500
Epoch 9/500
Epoch 10/500
Epoch 11/500
```

y_train = []

n_future = 1 # 1 future day

```
Epoch 12/500
Epoch 13/500
Epoch 14/500
Epoch 15/500
Epoch 16/500
Epoch 17/500
Epoch 18/500
Epoch 19/500
Epoch 20/500
Epoch 21/500
Epoch 22/500
Epoch 23/500
Epoch 24/500
Epoch 25/500
Epoch 26/500
Epoch 27/500
Epoch 28/500
Epoch 29/500
Epoch 30/500
Epoch 31/500
Epoch 32/500
Epoch 33/500
Epoch 34/500
Epoch 35/500
```

```
Epoch 36/500
Epoch 37/500
Epoch 38/500
Epoch 39/500
Epoch 40/500
Epoch 41/500
Epoch 42/500
Epoch 43/500
Epoch 44/500
Epoch 45/500
Epoch 46/500
Epoch 47/500
Epoch 48/500
Epoch 49/500
Epoch 50/500
Epoch 51/500
Epoch 52/500
Epoch 53/500
Epoch 54/500
Epoch 55/500
Epoch 56/500
Epoch 57/500
Epoch 58/500
Epoch 59/500
```

```
Epoch 60/500
Epoch 61/500
Epoch 62/500
Epoch 63/500
Epoch 64/500
Epoch 65/500
Epoch 66/500
Epoch 67/500
Epoch 68/500
Epoch 69/500
Epoch 70/500
Epoch 71/500
Epoch 72/500
Epoch 73/500
Epoch 74/500
Epoch 75/500
Epoch 76/500
Epoch 77/500
Epoch 78/500
Epoch 79/500
Epoch 80/500
Epoch 81/500
Epoch 82/500
Epoch 83/500
```

```
Epoch 84/500
Epoch 85/500
Epoch 86/500
Epoch 87/500
Epoch 88/500
Epoch 89/500
Epoch 90/500
Epoch 91/500
Epoch 92/500
Epoch 93/500
Epoch 94/500
Epoch 95/500
Epoch 96/500
Epoch 97/500
Epoch 98/500
Epoch 99/500
Epoch 100/500
Epoch 101/500
Epoch 102/500
Epoch 103/500
Epoch 104/500
Epoch 105/500
Epoch 106/500
Epoch 107/500
```

```
Epoch 108/500
Epoch 109/500
Epoch 110/500
Epoch 111/500
Epoch 112/500
Epoch 113/500
Epoch 114/500
Epoch 115/500
Epoch 116/500
Epoch 117/500
Epoch 118/500
Epoch 119/500
Epoch 120/500
Epoch 121/500
Epoch 122/500
Epoch 123/500
Epoch 124/500
Epoch 125/500
Epoch 126/500
Epoch 127/500
Epoch 128/500
Epoch 129/500
Epoch 130/500
Epoch 131/500
```

```
Epoch 132/500
Epoch 133/500
Epoch 134/500
Epoch 135/500
Epoch 136/500
Epoch 137/500
Epoch 138/500
Epoch 139/500
Epoch 140/500
Epoch 141/500
Epoch 142/500
Epoch 143/500
Epoch 144/500
Epoch 145/500
Epoch 146/500
Epoch 147/500
Epoch 148/500
Epoch 149/500
Epoch 150/500
Epoch 151/500
Epoch 152/500
Epoch 153/500
Epoch 154/500
Epoch 155/500
```

```
Epoch 156/500
Epoch 157/500
Epoch 158/500
Epoch 159/500
Epoch 160/500
Epoch 161/500
Epoch 162/500
Epoch 163/500
Epoch 164/500
Epoch 165/500
Epoch 166/500
Epoch 167/500
Epoch 168/500
Epoch 169/500
Epoch 170/500
Epoch 171/500
Epoch 172/500
Epoch 173/500
Epoch 174/500
Epoch 175/500
Epoch 176/500
Epoch 177/500
Epoch 178/500
Epoch 179/500
```

```
Epoch 180/500
Epoch 181/500
Epoch 182/500
Epoch 183/500
Epoch 184/500
Epoch 185/500
Epoch 186/500
Epoch 187/500
Epoch 188/500
Epoch 189/500
Epoch 190/500
Epoch 191/500
Epoch 192/500
Epoch 193/500
Epoch 194/500
Epoch 195/500
Epoch 196/500
Epoch 197/500
Epoch 198/500
Epoch 199/500
Epoch 200/500
Epoch 201/500
Epoch 202/500
Epoch 203/500
```

```
Epoch 204/500
Epoch 205/500
Epoch 206/500
Epoch 207/500
Epoch 208/500
Epoch 209/500
Epoch 210/500
Epoch 211/500
Epoch 212/500
Epoch 213/500
Epoch 214/500
Epoch 215/500
Epoch 216/500
Epoch 217/500
Epoch 218/500
Epoch 219/500
Epoch 220/500
Epoch 221/500
Epoch 222/500
Epoch 223/500
Epoch 224/500
Epoch 225/500
Epoch 226/500
Epoch 227/500
```

```
Epoch 228/500
Epoch 229/500
Epoch 230/500
Epoch 231/500
Epoch 232/500
Epoch 233/500
Epoch 234/500
Epoch 235/500
Epoch 236/500
Epoch 237/500
Epoch 238/500
Epoch 239/500
Epoch 240/500
Epoch 241/500
Epoch 242/500
Epoch 243/500
Epoch 244/500
Epoch 245/500
Epoch 246/500
Epoch 247/500
Epoch 248/500
Epoch 249/500
Epoch 250/500
Epoch 251/500
```

```
Epoch 252/500
Epoch 253/500
Epoch 254/500
Epoch 255/500
Epoch 256/500
Epoch 257/500
Epoch 258/500
Epoch 259/500
Epoch 260/500
Epoch 261/500
Epoch 262/500
Epoch 263/500
Epoch 264/500
Epoch 265/500
Epoch 266/500
Epoch 267/500
Epoch 268/500
Epoch 269/500
Epoch 270/500
Epoch 271/500
Epoch 272/500
Epoch 273/500
Epoch 274/500
Epoch 275/500
```

```
Epoch 276/500
Epoch 277/500
Epoch 278/500
Epoch 279/500
Epoch 280/500
Epoch 281/500
Epoch 282/500
Epoch 283/500
Epoch 284/500
Epoch 285/500
Epoch 286/500
Epoch 287/500
Epoch 288/500
Epoch 289/500
Epoch 290/500
Epoch 291/500
Epoch 292/500
Epoch 293/500
Epoch 294/500
Epoch 295/500
Epoch 296/500
Epoch 297/500
Epoch 298/500
Epoch 299/500
```

```
Epoch 300/500
Epoch 301/500
Epoch 302/500
Epoch 303/500
Epoch 304/500
Epoch 305/500
Epoch 306/500
Epoch 307/500
Epoch 308/500
Epoch 309/500
Epoch 310/500
Epoch 311/500
Epoch 312/500
Epoch 313/500
Epoch 314/500
Epoch 315/500
Epoch 316/500
Epoch 317/500
Epoch 318/500
Epoch 319/500
Epoch 320/500
Epoch 321/500
Epoch 322/500
Epoch 323/500
```

```
Epoch 324/500
Epoch 325/500
Epoch 326/500
Epoch 327/500
Epoch 328/500
Epoch 329/500
Epoch 330/500
Epoch 331/500
Epoch 332/500
Epoch 333/500
Epoch 334/500
Epoch 335/500
Epoch 336/500
Epoch 337/500
Epoch 338/500
Epoch 339/500
Epoch 340/500
Epoch 341/500
Epoch 342/500
Epoch 343/500
Epoch 344/500
Epoch 345/500
Epoch 346/500
Epoch 347/500
```

```
Epoch 348/500
Epoch 349/500
Epoch 350/500
Epoch 351/500
Epoch 352/500
Epoch 353/500
Epoch 354/500
Epoch 355/500
Epoch 356/500
Epoch 357/500
Epoch 358/500
Epoch 359/500
Epoch 360/500
Epoch 361/500
Epoch 362/500
Epoch 363/500
Epoch 364/500
Epoch 365/500
Epoch 366/500
Epoch 367/500
Epoch 368/500
Epoch 369/500
Epoch 370/500
Epoch 371/500
```

```
Epoch 372/500
Epoch 373/500
Epoch 374/500
Epoch 375/500
Epoch 376/500
Epoch 377/500
Epoch 378/500
Epoch 379/500
Epoch 380/500
Epoch 381/500
Epoch 382/500
Epoch 383/500
Epoch 384/500
Epoch 385/500
Epoch 386/500
Epoch 387/500
Epoch 388/500
Epoch 389/500
Epoch 390/500
Epoch 391/500
Epoch 392/500
Epoch 393/500
Epoch 394/500
Epoch 395/500
```

```
Epoch 396/500
Epoch 397/500
Epoch 398/500
Epoch 399/500
Epoch 400/500
Epoch 401/500
Epoch 402/500
Epoch 403/500
Epoch 404/500
Epoch 405/500
Epoch 406/500
Epoch 407/500
Epoch 408/500
Epoch 409/500
Epoch 410/500
Epoch 411/500
Epoch 412/500
Epoch 413/500
Epoch 414/500
Epoch 415/500
Epoch 416/500
Epoch 417/500
Epoch 418/500
Epoch 419/500
```

```
Epoch 420/500
Epoch 421/500
Epoch 422/500
Epoch 423/500
Epoch 424/500
Epoch 425/500
Epoch 426/500
Epoch 427/500
Epoch 428/500
Epoch 429/500
Epoch 430/500
Epoch 431/500
Epoch 432/500
Epoch 433/500
Epoch 434/500
Epoch 435/500
Epoch 436/500
Epoch 437/500
Epoch 438/500
Epoch 439/500
Epoch 440/500
Epoch 441/500
Epoch 442/500
Epoch 443/500
```

```
Epoch 444/500
Epoch 445/500
Epoch 446/500
Epoch 447/500
Epoch 448/500
Epoch 449/500
Epoch 450/500
Epoch 451/500
Epoch 452/500
Epoch 453/500
Epoch 454/500
Epoch 455/500
Epoch 456/500
Epoch 457/500
Epoch 458/500
Epoch 459/500
Epoch 460/500
Epoch 461/500
Epoch 462/500
Epoch 463/500
Epoch 464/500
Epoch 465/500
Epoch 466/500
Epoch 467/500
```

```
Epoch 468/500
Epoch 469/500
Epoch 470/500
Epoch 471/500
Epoch 472/500
Epoch 473/500
Epoch 474/500
Epoch 475/500
Epoch 476/500
Epoch 477/500
Epoch 478/500
Epoch 479/500
Epoch 480/500
Epoch 481/500
Epoch 482/500
Epoch 483/500
Epoch 484/500
Epoch 485/500
Epoch 486/500
Epoch 487/500
Epoch 488/500
Epoch 489/500
Epoch 490/500
Epoch 491/500
```

```
Epoch 492/500
Epoch 493/500
Epoch 494/500
Epoch 495/500
Epoch 496/500
Epoch 497/500
Epoch 498/500
Epoch 499/500
Epoch 500/500
```

Out [236]: <tensorflow.python.keras.callbacks.History at 0x13a2a2dfa60>

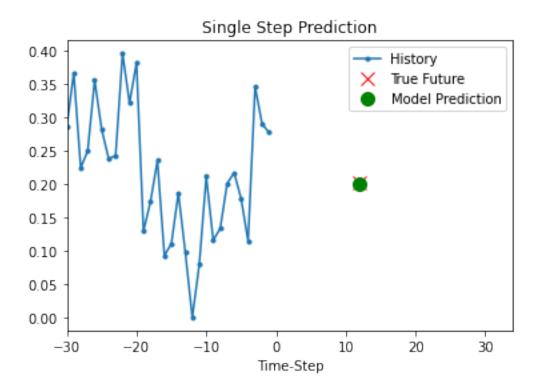
1.5.3 4.3 Test and Evaluate

return

1) Evaluate performance with single step prediction of Tempreture on the 31th day of 2019

```
In [276]: # Plot single step prediction
          # Modified from: Team, K. (n.d.). Keras documentation: Timeseries forecasting for we
          def show_plot(plot_data, delta, title):
              labels = ["History", "True Future", "Model Prediction"]
              marker = [".-", "rx", "go"]
              time_steps = list(range(-(plot_data[0].shape[0]), 0))
              if delta:
                  future = delta
              else:
                  future = 0
              plt.title(title)
              for i, val in enumerate(plot_data):
                      plt.plot(future, plot_data[i], marker[i], markersize=10, label=labels[i]
                  else:
                      plt.plot(time_steps, plot_data[i].flatten(), marker[i], label=labels[i])
              plt.legend()
              plt.xlim([time_steps[0], (future + 5) * 2])
              plt.xlabel("Time-Step")
              plt.show()
```

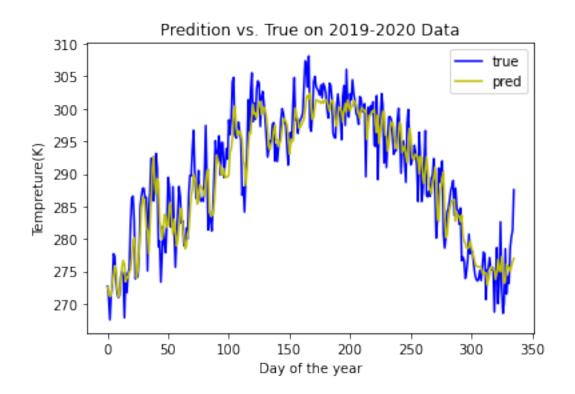
```
In [288]: # Preprocessing test data
          testdataset = cur_max_temp[:31]
          testdataset = sc.transform(testdataset)
          x_test = []
          y test = []
          n_future = 1
          n_past = 30
          for i in range(0,len(testdataset)-n_past-n_future+1):
              x_test.append(testdataset[i : i + n_past, 0])
              y_test.append(testdataset[i + n_past : i + n_past + n_future, 0])
          x_test , y_test = np.array(x_test), np.array(y_test)
          x_test = np.reshape(x_test, (x_test.shape[0] , x_test.shape[1], 1) )
In [299]: show_plot(
                  [x_test[0], y_test[0], regressor.predict(x_test)[0]],
                  12,
                  "Single Step Prediction",
              )
```



2) Evaluate performance with 2019-2020 test set

```
x_test = []
          y_test = []
          n_future = 1
          n_past = 30
          for i in range(0,len(testdataset)-n_past-n_future+1):
              x_test.append(testdataset[i : i + n_past, 0])
              y_test.append(testdataset[i + n_past : i + n_past + n_future, 0])
          x_test , y_test = np.array(x_test), np.array(y_test)
          x_test = np.reshape(x_test, (x_test.shape[0] , x_test.shape[1], 1) )
In [301]: # Making prediction and evaluation
          pred_30 = regressor.predict(x_test)
          pred_30 = sc.inverse_transform(pred_30)
          y_test = sc.inverse_transform(y_test)
          mse = mean_squared_error(y_test, pred_30)
          print('MSE for Tempreture Prediction using LSTM: ', mse)
          plt.plot(y_test, color = 'b', label = "true")
          plt.plot(pred_30, color = 'y', label = "pred")
          plt.ylabel('Tempreture(K)')
          plt.xlabel('Day of the year')
          plt.title("LSTM prediction on 2019-2020 Data")
          plt.legend()
          plt.show()
```

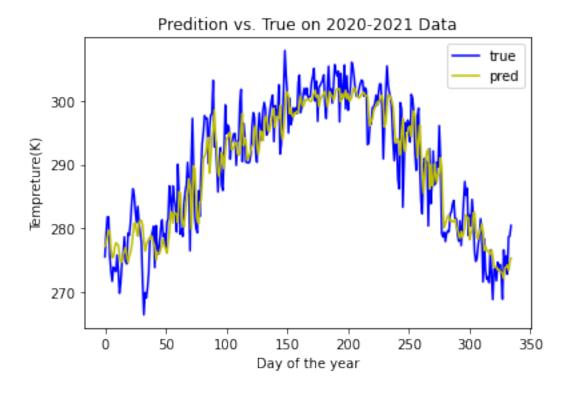
MSE for Tempreture Prediction using LSTM: 14.898463512340081



3) Evaluate performance with 2020-2021 test set

```
In [302]: # Preprocessing test set
          testdataset = future_max_temp
          testdataset = sc.transform(testdataset)
          x test = []
          y_test = []
          n_future = 1
          n_past = 30
          for i in range(0,len(testdataset)-n_past-n_future+1):
              x_test.append(testdataset[i : i + n_past, 0])
              y_test.append(testdataset[i + n_past : i + n_past + n_future, 0])
          x_test , y_test = np.array(x_test), np.array(y_test)
          x_test = np.reshape(x_test, (x_test.shape[0] , x_test.shape[1], 1) )
In [303]: # Making prediction and evaluation
          pred_2021 = regressor.predict(x_test)
          pred_2021 = sc.inverse_transform(pred_2021)
          y_test = sc.inverse_transform(y_test)
          mse = mean_squared_error(y_test, pred_2021)
          print('MSE for Tempreture Prediction using LSTM: ', mse)
          plt.plot(y_test, color = 'b', label = "true")
          plt.plot(pred 2021, color = 'y', label = "pred")
          plt.ylabel('Tempreture(K)')
          plt.xlabel('Day of the year')
          plt.title("LSTM prediction on 2020-2021 Data")
          plt.legend()
          plt.show()
```

MSE for Tempreture Prediction using LSTM: 16.6580981560278



1.6 References:

Kosandal, R. (2020, January 5). Weather forecasting with Recurrent Neural Networks. Medium. https://medium.com/analytics-vidhya/weather-forecasting-with-recurrent-neural-networks-1eaa057d70c3. Team, K. (n.d.). Keras documentation: Timeseries forecasting for weather https://keras.io/examples/timeseries/timeseries_weather_forecasting/. prediction. Fatmakursun. (2020, January 28). Rain Forecasting with Artificial Neural Network. Kaggle. https://www.kaggle.com/fatmakursun/rain-forecasting-with-artificial-neural-network. Pietro, M. D. (2020, December 8). Machine Learning with Python: Regression. Medium. https://towardsdatascience.com/machine-learning-with-python-regression-complete-tutorial-47268e546cea. Paialunga, P. (2021, April 18). Weather forecasting with Machine Learning, using Python. Medium. https://towardsdatascience.com/weather-forecasting-with-machinelearning-using-python-55e90c346647. Gabriel, L. (2019, April 29). A Beginner's Guide on Recurrent Neural Networks with PyTorch. FloydHub. https://blog.floydhub.com/a-beginnersguide-on-recurrent-neural-networks-with-pytorch/ George, S. (2019, May 20). Understanding the 3 most common loss functions for Machine Learning Regression. TowardsDataScience. https://towardsdatascience.com/understanding-the-3-most-common-loss-functions-formachine-learning-regression-23e0ef3e14d3 Part of the Recurrent Neural Network structure code credits to Kristina's project 3 and 4 of CS4740, Fall 2020.