Tomorrow's Weather, Today!

Weather Prediction using Machine Learning

Patrick Sigourney
Data Mining - SP 2019

Goal

Accurately predict weather for the next 3 days based on actual weather data from the previous 5 days.

Data

Powered by Dark Sky

DarkSky.net API

10 years of historic daily weather data at UT

Metrics Used:

- High Temp
- Low Temp
- Precipitation Probability
- Barometric Pressure
- Cloud Cover Percentage
- Wind Speed
- Wind Direction

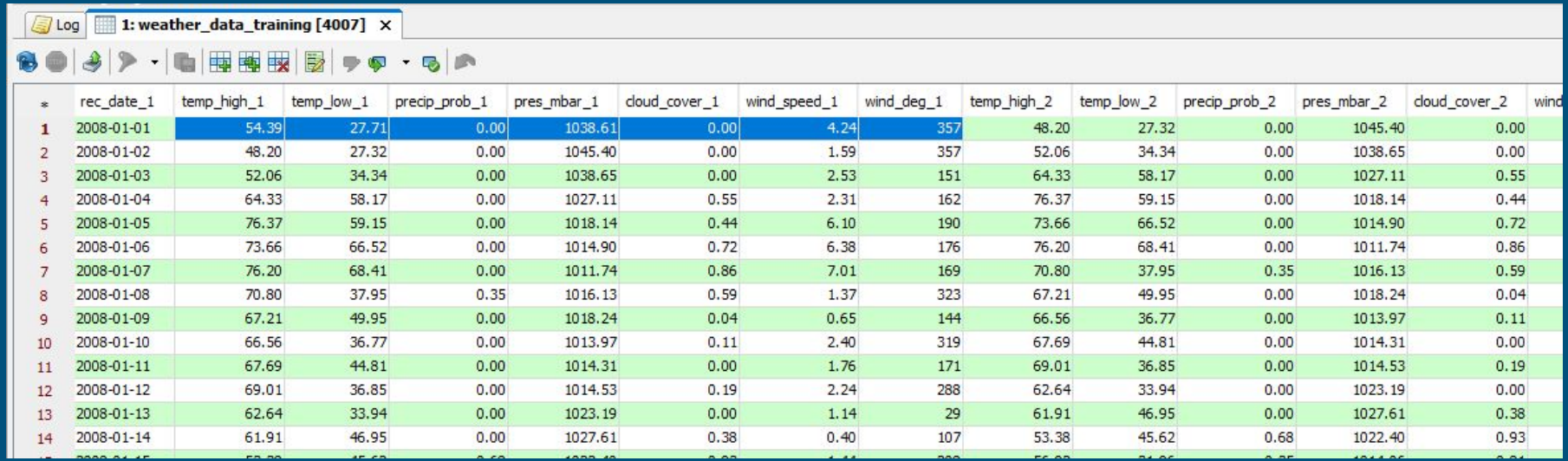
Models

One model for each predicted feature

Ridge regression with optimized alpha

Training

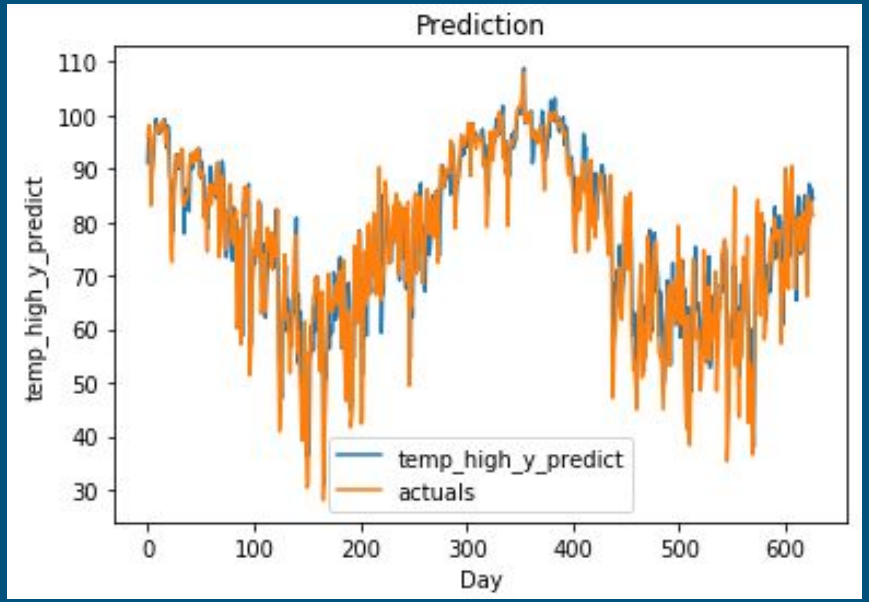
Training data arranged into records of 5+1 days
Models trained on 2008-2018 data; 2019 data used for testing

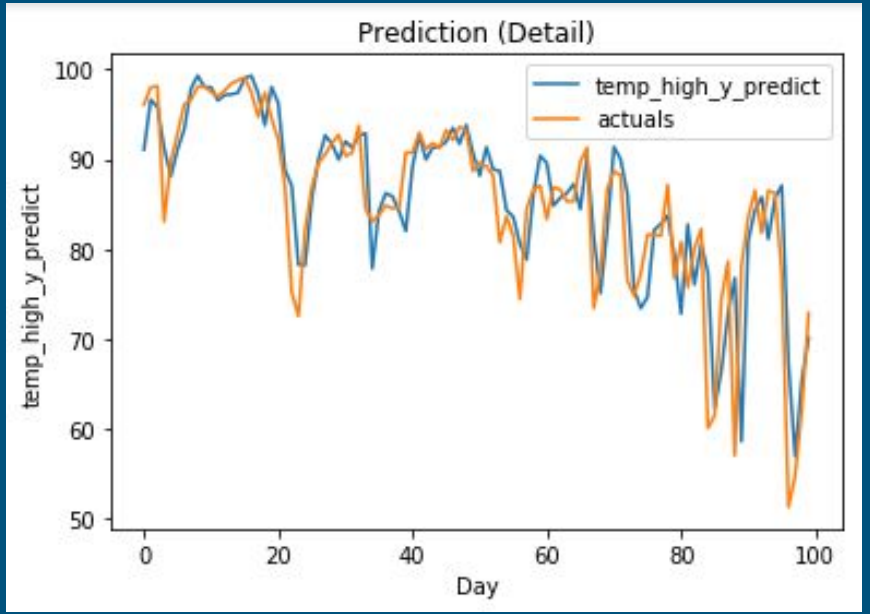


Log | 1: weather_data_training [4007]

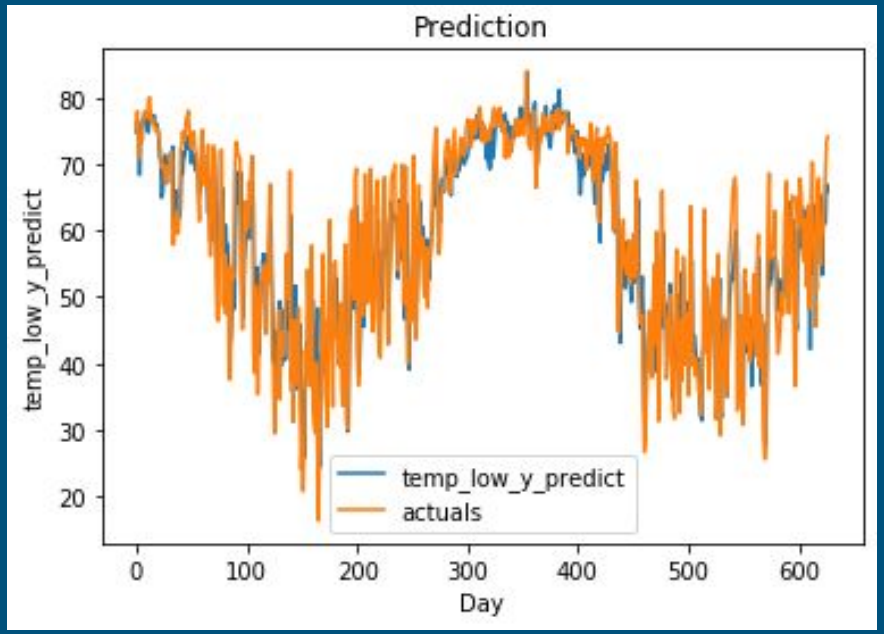
*	rec_date_1	temp_high_1	temp_low_1	precip_prob_1	pres_mbar_1	cloud_cover_1	wind_speed_1	wind_deg_1	temp_high_2	temp_low_2	precip_prob_2	pres_mbar_2	cloud_cover_2	wind
1	2008-01-01	54.39	27.71	0.00	1038.61	0.00	4.24	357	48.20	27.32	0.00	1045.40	0.00	
2	2008-01-02	48.20	27.32	0.00	1045.40	0.00	1.59	357	52.06	34.34	0.00	1038.65	0.00	
3	2008-01-03	52.06	34.34	0.00	1038.65	0.00	2.53	151	64.33	58.17	0.00	1027.11	0.55	
4	2008-01-04	64.33	58.17	0.00	1027.11	0.55	2.31	162	76.37	59.15	0.00	1018.14	0.44	
5	2008-01-05	76.37	59.15	0.00	1018.14	0.44	6.10	190	73.66	66.52	0.00	1014.90	0.72	
6	2008-01-06	73.66	66.52	0.00	1014.90	0.72	6.38	176	76.20	68.41	0.00	1011.74	0.86	
7	2008-01-07	76.20	68.41	0.00	1011.74	0.86	7.01	169	70.80	37.95	0.35	1016.13	0.59	
8	2008-01-08	70.80	37.95	0.35	1016.13	0.59	1.37	323	67.21	49.95	0.00	1018.24	0.04	
9	2008-01-09	67.21	49.95	0.00	1018.24	0.04	0.65	144	66.56	36.77	0.00	1013.97	0.11	
10	2008-01-10	66.56	36.77	0.00	1013.97	0.11	2.40	319	67.69	44.81	0.00	1014.31	0.00	
11	2008-01-11	67.69	44.81	0.00	1014.31	0.00	1.76	171	69.01	36.85	0.00	1014.53	0.19	
12	2008-01-12	69.01	36.85	0.00	1014.53	0.19	2.24	288	62.64	33.94	0.00	1023.19	0.00	
13	2008-01-13	62.64	33.94	0.00	1023.19	0.00	1.14	29	61.91	46.95	0.00	1027.61	0.38	
14	2008-01-14	61.91	46.95	0.00	1027.61	0.38	0.40	107	53.38	45.62	0.68	1022.40	0.93	

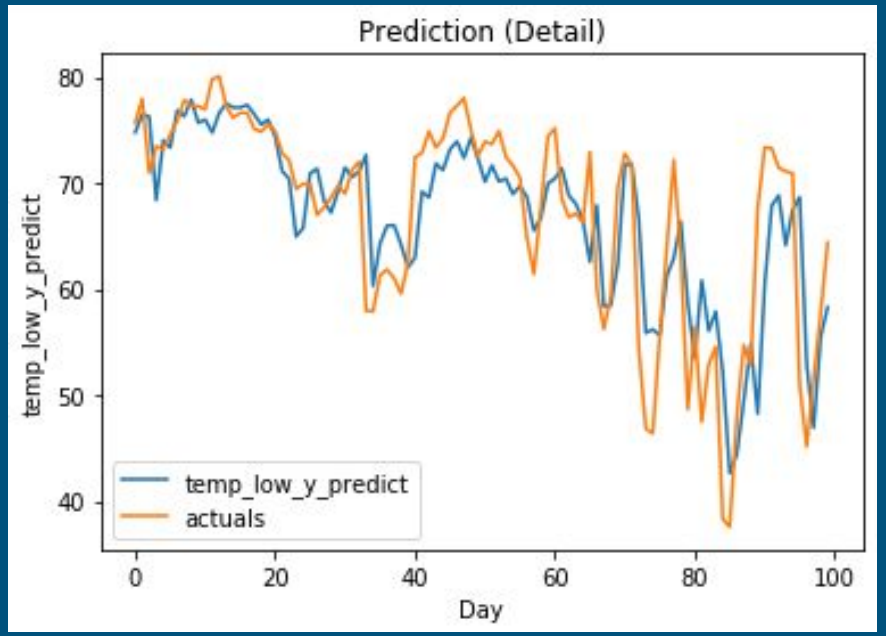
Results


Prediction
temp_high_y_predict
Day
temp_high_y_predict
actuals
110
100
90
80
70
60
50
40
30
0
100
200
300
400
500
600

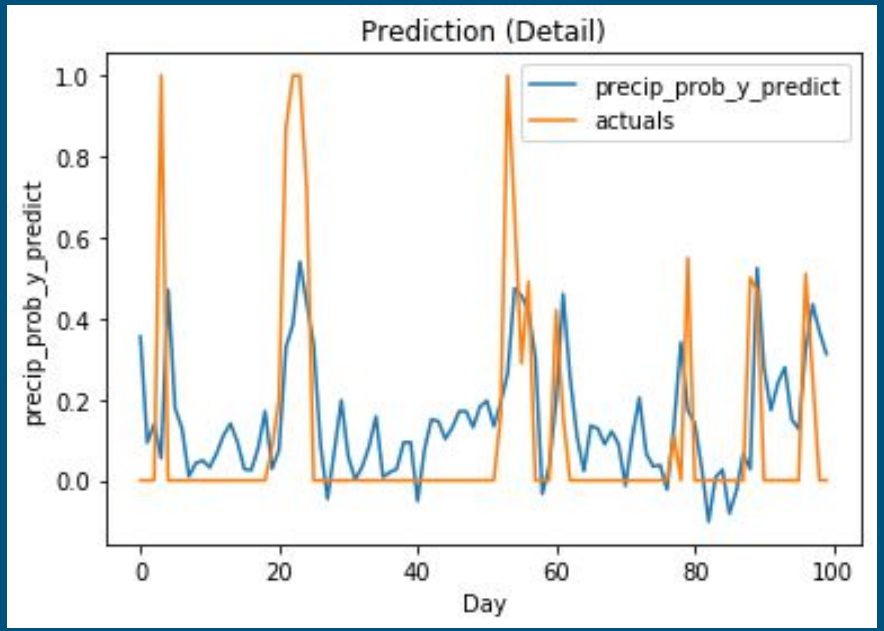

Prediction (Detail)
temp_high_y_predict
Day
temp_high_y_predict
actuals
100
90
80
70
60
50
0
20
40
60
80
100

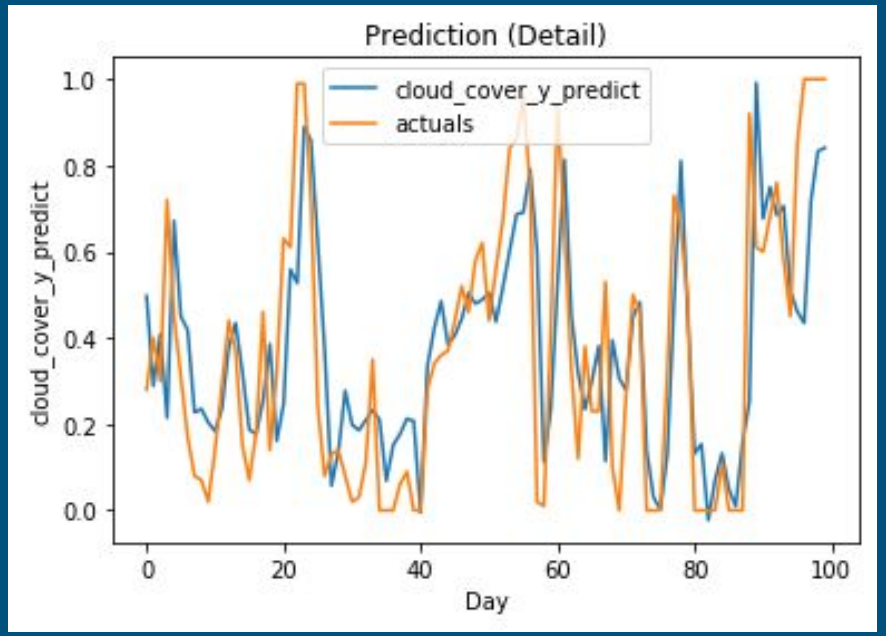
Results


Prediction
temp_low_y_predict
Day
temp_low_y_predict
actuals
20
30
40
50
60
70
80
0
100
200
300
400
500
600


Prediction (Detail)
temp_low_y_predict
Day
temp_low_y_predict
actuals
40
50
60
70
80
0
20
40
60
80
100

Results


Prediction (Detail)
precip_prob_y_predict
actuals
precip_prob_y_predict
Day
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100


Prediction (Detail)
cloud_cover_y_predict
actuals
cloud_cover_y_predict
Day
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100

Relevant Features

For high temp prediction, the most relevant features were the cloud cover measurements of the previous 2 days.

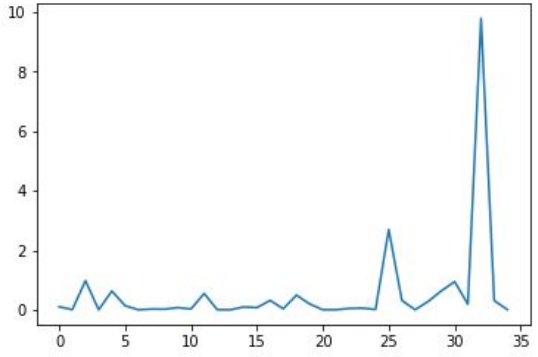
Mean absolute error: 6.73 degrees
Mean absolute deviation: 5.25 degrees

In [22]:
```
##High Temp Coefficients

import matplotlib.pyplot as plt
%matplotlib inline

coef = model.coef_

plt.plot(abs(coef))
plt.show()
```



Relevant Features

For low temp prediction, the most relevant features were the cloud cover and precipitation probability measurements from the previous day and 3 days prior.

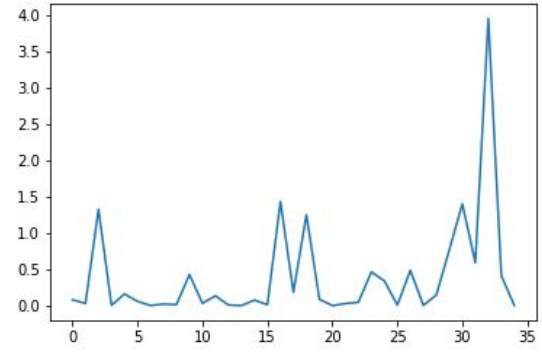
Mean absolute error: 5.94 degrees
Mean absolute deviation: 5.0 degrees

```
In [20]: ##Low Temp Coefficients

         import matplotlib.pyplot as plt
         %matplotlib inline

         coef = model.coef_

         plt.plot(abs(coef))
         plt.show()
```



Relevant Features

For precipitation probability prediction, the most relevant features were the cloud cover and precipitation probability measurements from the previous day the cloud cover from two days prior.

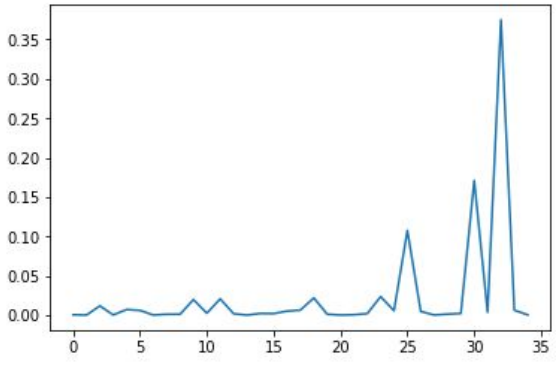
Precip Probability mean abs error: 0.17
Precip Probability mean abs dev: 0.14
Cloud cover mean abs error: 0.27
Cloud cover mean abs dev: 0.22

In [25]:
```
##Precip Prob Coefficients

import matplotlib.pyplot as plt
%matplotlib inline

coef = model.coef_

plt.plot(abs(coef))
plt.show()
```



Prediction

Multi-day predictions require using single-day predicted values as inputs.

The further out the prediction, the less accurate it becomes.

We saw the most relevant feature in all the predictions is the previous day's cloud cover and precipitation probability, but the amount of error in these predictions is so high it suggests the multi-day predictions which rely on these values will be highly inaccurate.

Prediction

High Temp:

1st Day Mean Abs Error: 6.73 degrees; mean abs deviation: 5.25 degrees
2nd Day Mean Abs Error: 9.32 degrees; mean abs deviation: 7.38 degrees
3rd Day Mean Abs Error: 10.5 degrees; mean abs deviation: 8.03 degrees

Low Temp:

1st Day Mean Abs Error: 5.94 degrees; mean abs deviation: 5.0 degrees
2nd Day Mean Abs Error: 9.07 degrees; mean abs deviation: 6.08 degrees
3rd Day Mean Abs Error: 10.05 degrees; mean abs deviation: 7.0 degrees

Prediction

Cloud Coverage:

1st Day Mean Abs Error: 0.27; mean abs deviation: 0.22
2nd Day Mean Abs Error: 0.33; mean abs deviation: 0.19
3rd Day Mean Abs Error: 0.33; mean abs deviation: 0.17

Precipitation Probability:

1st Day Mean Abs Error: 0.17; mean abs deviation: 0.14
2nd Day Mean Abs Error: 0.18; mean abs deviation: 0.12
3rd Day Mean Abs Error: 0.18; mean abs deviation: 0.06

Demo