



菊池研発表課題

2-3-29

高松毅瑠



天気を分析しよう

- 気象庁からデータを取ってきてそのデータからグラフの作成、Rでやった決定木の作製などをやってみる。

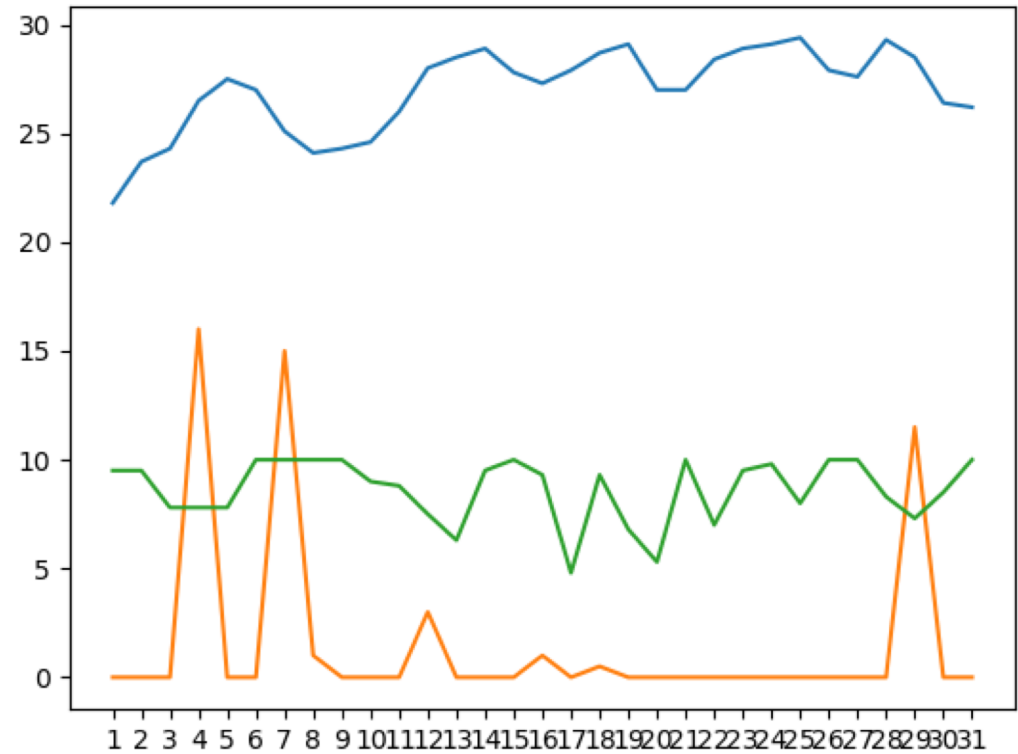


元となるデータ

- ▶ 年月日,平均気温,降水量の合計,合計全天日射量,平均風速,平均湿度,平均雲量,天気概況(昼:06時~18時)
- ▶ (例)2008/7/1 21.8 0 20.55 2.7 71 9.5 薄曇後一時晴
- ▶ 7/1~7/31のデータを2008~2018まで11年分×31日分集めた。

出力してみた

- 青-気温
- オレンジ-降水量
- 緑-雲量

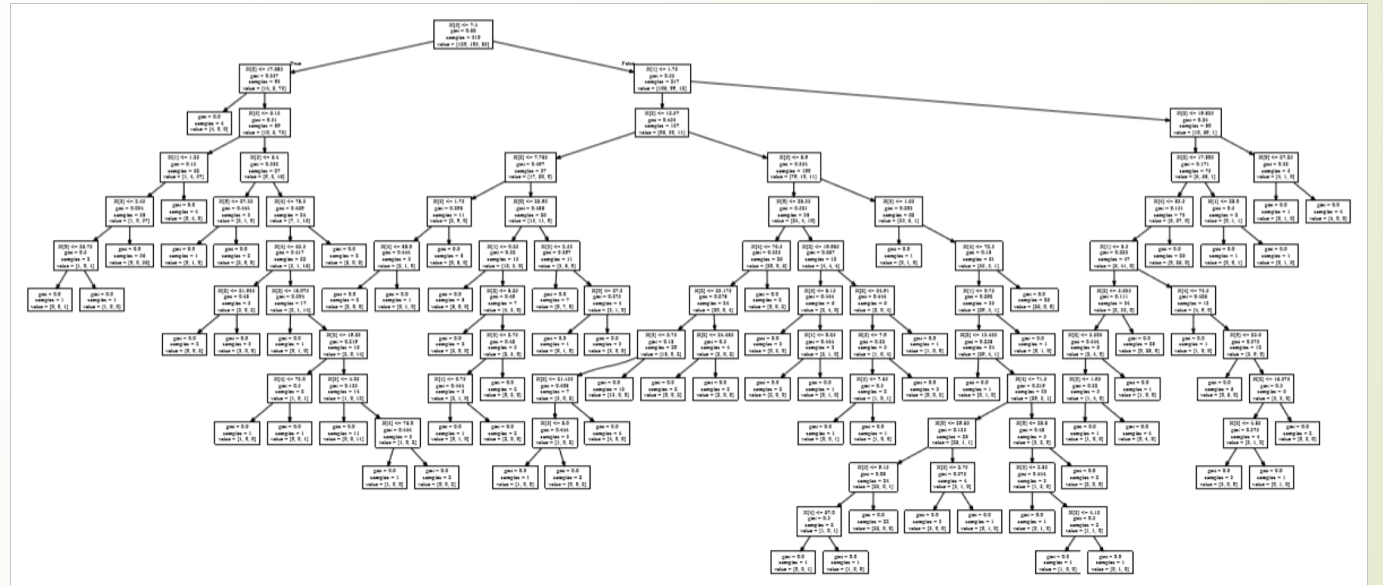


Rと同じようなことをしてみる

- ▶ 目的変数を天気、それ以外の気温、降水量などを説明変数としてその日の天気を予測してみる。
- ▶ また薄雲後一時晴れなどのあいまいな表現に対しては最初に雲、薄雲があれば曇り、晴れ、快晴がある場合は晴れとし、雨という単語があった場合は雨にした。
- ▶ 2008~2017の10年分のデータを学習させ、2018のデータから天気を予測してみる。

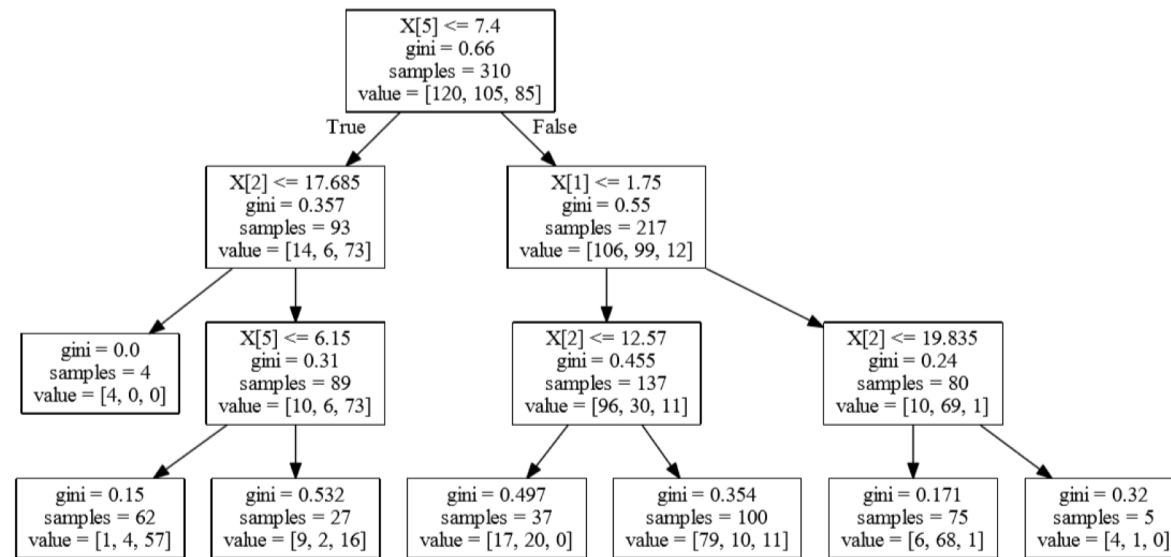
結果

- ▶ なにも設定しないとこうなる。
- ▶ また結果は0.7666666666666667となり約77パーセントしか予測が合わなかった。



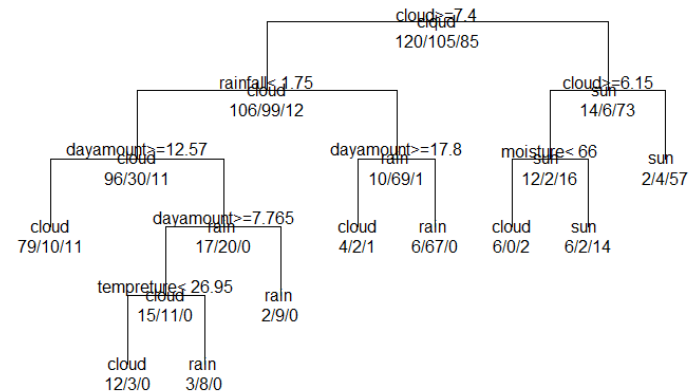
より結果が良くなるように...


- ▶ Maxdepthを3に設定することで単純な木になった。また結果も0.9となり90%を予測できたのであるからこのほうが良い



Rで同じことをしてみる

- ▶ Rを使ってやってみた。データの修正はpythonを使って行ったPythonは深さなどの設定をしなければならぬがRだと勝手に設定されていた。
- ▶ また結果はほとんど同じになった。





考察

- 今までRで使ったデータはあらかじめ形が整えられていてすぐに利用することができたが、利用するデータがそうでないデータの時に求めるデータの形に修正できるので便利だと思った。
- その日のデータではなく前日のデータや気象情報から予測してみることもできると思った。