Individual Machine Learning Project (Australian Weather Prediction)

Task

Predict next-day rain by training classification model (Logistic Regression) on the target variable RainTomorrow. Use your Jupyter notebook for this.

You will be graded on the following:

- Your ability to import the data into Jupyter notebook, all needed packages, and use various Pandas methods like head, info, etc. to get explore the contents of the data.
- How you treat missing values
- How you choose which features to use to train the model
- How you treat categorical features like location, etc.
- How you split data into training and test sets for the model.
- How the model is trained
- How you evaluate model prediction results
- Layout of notebook and comments on code and code outputs

The dataset

The dataset contains about 10 years of daily weather observations from many locations across Australia.

RainTomorrow is the target variable to predict. It means -- did it rain the next day, Yes or No? This column is Yes if the rain for that day was 1mm or more.

Date: the date of the observation

Location: The common name of the location of the weather station.

MinTemp: The minimum temperature in degree Celsius

MaxTemp: The maximum temperature in degree Celsius

Rainfall: The amount of rainfall recorded for the day in mm

Evaporation: The so-called Class A pan evaporation(mm) in the 25 hours to 9am

Sunshine: The number of hours of bright sunshine in the day

WindGustDir: The direction of the strongest wind gust in the 24 hours to midnight

WindGustSpeed: The speed (km/h) of the strongest wind gust in the 24 hours to midnight

WindDir9am: Direction of the wind at 9am

WindDir3pm: Direction of the wind at 3pm

WindSpeed9am: Wind speed (km/hr) averaged over 10 minutes prior to 9am

WindSpeed3pm: Wind speed (km/hr) averaged over 10 minutes prior to 3pm

Humidity9am: Humidity (percent) at 9am

Humidity3pm: Humidity (percent) at 3pm

Pressure9am: Atmospheric pressure (hpa) reduced to mean sea level at 9am

Pressure3pm: Atmospheric pressure(hpa) reduced to mean sea level at 3pm

Cloud9am: Fraction of the sky obscured by cloud at 9am. This is measured in “oktas” which are unit of eighths.

Cloud3pm: Fraction of the sky obscured by cloud at 3pm. This is measured in “oktas” which are unit of eighths.

Temp9am: Temperature (degrees C) at 9am

Temp3pm: Temperature (degrees C) at 3pm

RainToday: Yes, if precipitation(mm) in the 24 hours to 9am exceeds 1mm, otherwise No

Source & Acknowledgements

Observations were drawn from numerous weather stations. The daily observations are available from http://www.bom.gov.au/climate/data.
An example of latest weather observations in Canberra:
http://www.bom.gov.au/climate/dwo/IDCJDW2801.latest.shtml

Definitions adapted from http://www.bom.gov.au/climate/dwo/IDCJDW0000.shtml
Data source: http://www.bom.gov.au/climate/dwo/ and http://www.bom.gov.au/climate/data.