**103.El Nino Data Set（TODO：补充简介）**

1. 数据库网址

https://archive.ics.uci.edu/ml/datasets/El+Nino

2. 数据库描述

【1.[数据集名称]数据集由[机构名或人名]采集；】The data used in our experiments were collected by E. Alpaydin, C. Kaynak, from Department of Computer Engineering,Bogazici University at July,1998.【2.用于[什么实验目的]】We used preprocessing programs made available by NIST to extract normalized bitmaps of handwritten digits from a preprinted form.【3】

【4】The database has 5620 samples, respectively belong to optdigits.tra with 3823 samples and optidigits.tes with 1797 samples. The categories of network system include seven categories, as shown in Table 1.

Table 1 Category Distribution of Network System [根据数据库绘制]

|  |  |  |  |
| --- | --- | --- | --- |
| Invasion Categories | optdigits.tra | optdigits.tes | Total Number of Samples |
|  |  |  |  |
|  |  |  |  |
| Total number of samples in total |  |  | 178080 |

|  |  |
| --- | --- |
| **Abstract**: The data set contains oceanographic and surface meteorological readings taken from a series of buoys positioned throughout the equatorial Pacific. |  |

**Source:**

Original Owner:   
  
Pacific Marine Environmental Laboratory   
National Oceanic and Atmospheric Administration   
US Department of Commerce   
<http://www.pmel.noaa.gov/>   
  
Donor:   
  
Dr Di Cook   
Department of Statistics   
Iowa State University   
dicook **'@'** iastate.edu   
<http://www.public.iastate.edu/~dicook/>

**Data Set Information:**

This data was collected with the Tropical Atmosphere Ocean (TAO) array which was developed by the international Tropical Ocean Global Atmosphere (TOGA) program. The TAO array consists of nearly 70 moored buoys spanning the equatorial Pacific, measuring oceanographic and surface meteorological variables critical for improved detection, understanding and prediction of seasonal-to-interannual climate variations originating in the tropics, most notably those related to the El Nino/Southern Oscillation (ENSO) cycles.   
  
The moorings were developed by National Oceanic and Atmospheric Administration's (NOAA) Pacific Marine Environmental Laboratory (PMEL). Each mooring measures air temperature, relative humidity, surface winds, sea surface temperatures and subsurface temperatures down to a depth of 500 meters and a few a of the buoys measure currents, rainfall and solar radiation. The data from the array, and current updates, can be viewed on the web at the this address .   
  
The El Nino/Southern Oscillation (ENSO) cycle of 1982-1983, the strongest of the century, created many problems throughout the world. Parts of the world such as Peru and the Unites States experienced destructive flooding from increased rainfalls while the western Pacific areas experienced drought and devastating brush fires. The ENSO cycle was neither predicted nor detected until it was near its peak. This highlighted the need for an ocean observing system (i.e. the TAO array) to support studies of large scale ocean-atmosphere interactions on seasonal-to-interannual time scales.   
  
The TAO array provides real-time data to climate researchers, weather prediction centers and scientists around the world. Forcasts for tropical Pacific Ocean temperatures for one to two years in advance can be made using the ENSO cycle data. These forcasts are possible because of the moored buoys, along with drifting buoys, volunteer ship temperature probes, and sea level measurements.   
  
Research questions of interest include:   
  
- How can the data be used to predict weather conditions throughout the world?   
- How do the variables relate to each other?   
- Which variables have a greater effect on the climate variations?   
- Does the amount of movement of the buoy effect the reliability of the data?   
- When performing an analysis of the data, one should pay attention the possible affect of autocorrelation. Using a multiple regression approach to model the data would require a look at autoregression since the weather statistics of the previous days will affect today's weather.   
  
The data is stored in an ASCII files with one observation per line. Spaces separate fields and periods (.) denote missing values.   
  
More information and data from the TAO array can be found at the Pacific Marine Environmental Laboratory TAO data webpage: [[Web Link]](http://www.pmel.noaa.gov/toga-tao/)   
  
Information on storm data is available here: [[Web Link]](http://www.ncdc.noaa.gov/pdfs/sd/sd.html). This site contains data from January 1994 to April 1998 in a chronological listing by state provided by the National Weather Service. The data includes hurricanes, tornadoes, thunderstorms, hail, floods, drought conditions, lightning, high winds, snow, and temperature extremes.   
  
Hurricane tracking data for the Atlantic is available here: [[Web Link]](http://wxp.eas.purdue.edu/hur_atlantic/). The site contains a map showing the paths of the Atlantic hurricanes and also includes the storms winds (in knots), pressure (in millibars), and the category of the storm based on Saffir-Simpson scale.   
  
Another site of interest related to the ENSO cyles is available here: [[Web Link]](http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/current_impacts/precip_accum.html). This site contains information on twelve areas of the world that have demonstrated ENSO-precipitation relationships. Included in the site are maps of the areas and time series plots of actual daily precipitation and accumulated normal precipitation for the areas.

**Attribute Information:**

The data consists of the following variables: date, latitude, longitude, zonal winds (west<0, east>0), meridional winds (south<0, north>0), relative humidity, air temperature, sea surface temperature and subsurface temperatures down to a depth of 500 meters. Data taken from the buoys from as early as 1980 for some locations. Other data that was taken in various locations are rainfall, solar radiation, current levels, and subsurface temperatures.   
  
The latitude and longitude in the data showed that the bouys moved around to different locations. The latitude values stayed within a degree from the approximate location. Yet the longitude values were sometimes as far as five degrees off of the approximate location.   
  
Looking at the wind data, both the zonal and meridional winds fluctuated between -10 m/s and 10 m/s. The plot of the two wind variables showed no linear relationship. Also, the plots of each wind variable against the other three meteorolgical data showed no linear relationships.   
  
The relative humidity values in the tropical Pacific were typically between 70% and 90%.   
  
Both the air temperature and the sea surface temperature fluctuated between 20 and 30 degrees Celcius. The plot of the two temperatures variables shows a positive linear relationship existing. The two temperatures when each plotted against time also have similar plot designs. Plots of the other meteorological variables against the temperature variables showed no linear relationship.   
  
There are missing values in the data. As mentioned earlier, not all buoys are able to measure currents, rainfall, and solar radiation, so these values are missing dependent on the individual buoy. The amount of data available is also dependent on the buoy, as certain buoys were commissioned earlier than others.   
  
All readings were taken at the same time of day.