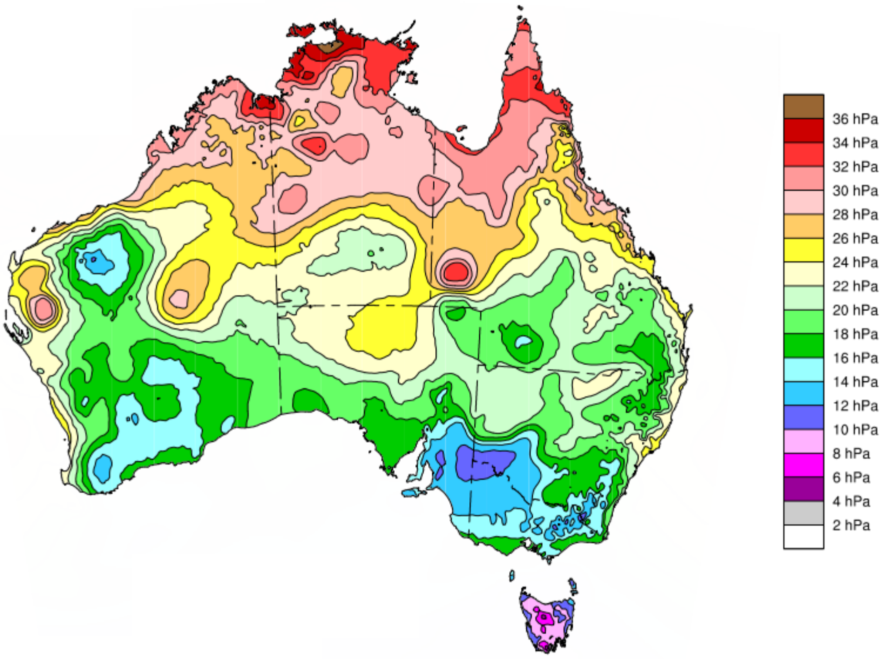
HW-1：Linear Regression

Deadlines: **2016.04.05**-23:59:59

In this homework, you are asked to rebuild the **vapour pressure** map of Australia based on the training data. The following three approaches need to be realized respectively:

- **Maximum likelihood approach (ML)**
- **Maximum a posteriori approach (MAP)**
- **Bayesian approach**

Besides, the cross-validation is suggested for you to select the model parameters for the surface fitting problem.



Data

Input data are the measured **vapour pressure** across Australia. Two data files are offered: **Training _data_hw1.mat** file and one **Test_data1_hw1.mat** file.

Training _data_hw1.mat:

- **X_train** is a 50000x3 matrix, whose first column and second column record the vertical position x_1 and horizontal position x_2 of the 50000 training points, respectively. The third column x_3 records the "**sea index**". For each data point, the corresponding entry $x_3 = 1$ for sea and $x_3 = 0$ for land.

- **T_train is** a 50000x1 vector, which records the vapour pressure (target values) of the training points.
- **x1_bound** is a vector recording the min/max values of the vertical index.
- **x2_bound** is a vector recording the min/max values of the horizontal index.

Test_data1_hw1.mat:

- **X_test** is a 10000x3 matrix, whose first column and second column record the vertical position and horizontal position of 10000 test points, respectively. The third column records the "**sea index**".
- **T_test** is a 10000x1 vector, which records the vapour pressure (target values) of the test points.

X_all_data.mat:

- **X_all** is a 612262x3 matrix, whose first column and second column record the vertical position and horizontal position of 612262 test points, respectively. The third column records the "**sea index**". <u>**These data points cover all the pixels on the map.**</u>

◆ Model

In this homework, your model should be implemented by the <u>**feature vector**</u>, which is defined as

$$\boldsymbol{\phi}(\mathrm{x}) = [\phi_1(x), \phi_2(x), \dots, \phi_N(x), \phi_{bias}(x), \phi_{sea}(x)\ ]$$

Using the linear combination of feature functions to predict the data

$$y(\boldsymbol{x}, \boldsymbol{w}) = \sum_{n}^{N+2} w_n \phi_N(\boldsymbol{x})$$

and minimizing the mean square error function

$$E(\boldsymbol{w}) = \frac{1}{2K} \sum_{k}^{K} ||y(\boldsymbol{x}(k), \boldsymbol{w}) - t(k)||^2$$

For example, we uniformly place N Gaussian basis functions over the spatial domain, with $N = O_1 \times O_2$. Here, O_1 and O_2 denote the number of locations along the horizontal and vertical directions, respectively.

For $1 \le n \le N$, we define the Gaussian basis functions as

$$\phi_n(x) = \exp\left(-\frac{(x_1-\mu_i)^2}{2s_{1n}^2} - \frac{\left(x_2-\mu_j\right)^2}{2s_{2n}^2}\right), for\ 1 \le i \le O_1, 1 \le j \le O_2$$

where $$n = \mathrm{O}_2 \times (i-1) + j.$$

Beside the N Gaussian basis function, another two components $\phi_{bias}(x)$ and $\phi_{sea}(x)$ can also be used for the regression problem.

$$\phi_{bias}(x) = c$$

$$\phi_{sea}(x) = x_{sea-index}$$

The example is only for reference, which means your model doesn't have to be the same! Please design the appropriate feature vector to fit the training data as well as possible.

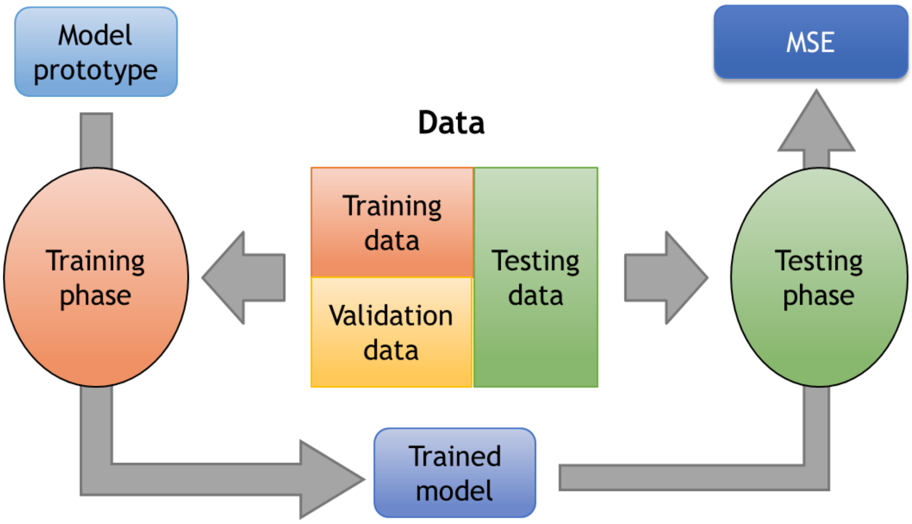
◆ Tasks

1. ML approach/MAP approach/Bayesian approach

Use the models you design based on three different approaches to predict the vapour pressure for each point in **X_test**. Plot the predicted vapour pressure map and the squared error distribution. Please discuss and explain the differences among these three approaches in detail.

2. Cross validation

Revisit the Bayesian approach. Apply N-fold cross-validation in your training stage to select the best mode parameters.



Requirements for demonstration:

- ☐ The **predicted vapour pressure map** for every approach (691x886) which is contained in a **.mat** file. This should be how your maps look like：

$$\begin{bmatrix} \hat{y}(10.00,112.00) & \hat{y}(10.00,112.05) & \cdots & \hat{y}(10.00,156.20) & \hat{y}(10.00,156.25) \\ \hat{y}(10.05,112.00) & \hat{y}(10.05,112.05) & \cdots & \hat{y}(10.05,156.20) & \hat{y}(10.05,156.25) \\ \vdots & & \ddots & & \vdots \\ \hat{y}(44.45,112.00) & \hat{y}(44.45,112.05) & \cdots & \hat{y}(44.45,156.20) & \hat{y}(44.45,156.25) \\ \hat{y}(44.50,112.00) & \hat{y}(44.50,112.05) & \cdots & \hat{y}(44.50,156.20) & \hat{y}(44.50,156.25) \end{bmatrix}$$

where $\hat{y}(x_1, x_2) = y(x_1, x_2, \boldsymbol{w})$.

- ☐ **Explanation** of your model.
- ☐ Five minutes for each person.

Reminders:

- ☐ The **pressure data on the sea** is likewise important!
- ☐ Please make sure your source code can be compiled by **Matlab**.
- ☐ **DO NOT COPY!!!（懶人包、考古題亦同）**