EE141L: Lab 7
Least squares and linear regression

Nov. 2, 2018

1 Linear regression

For many applications there is a need to have tools to help predict future events based on past observations. Examples include:

1. Stock price forecasting: predict tomorrow's price of a certain stock, based on its price the previous days, and prices of other related stocks.
2. Wind speed for wind energy generation: predict wind velocity based on current and previous weather conditions, in order to estimate the amount of power a wind farm will be delivering to the electric grid.
3. State of charge of batteries: estimate the time until battery runs out. Obvious applications are phones, computers, and electric cars.
4. Weather: predict temperature, humidity, precipitation, etc.

Consider a sequence of values $x(1), x(2), \cdots$ that represent some quantity (e.g. temperature), where the value $x(t)$ is , for example, the temperature at time t. A simple approach for prediction is by assuming a linear model[1]:

$$x(t) = \beta_1 x(t-1) + \beta_2 x(t-2) + \cdots, +\beta_B(t-B) + e(t), \tag{1}$$

that is, the temperature at time t is a linear combination of the temperatures at times $t-1, t-2, \cdots, t-B$ plus some (hopefully small) error $e(t)$.

1.1 Finding $\boldsymbol{\beta}$

The vector $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_B]^\top$ is unknown, so the first part of this lab is dedicated to estimating $\boldsymbol{\beta}$ from data. Assume we have the data vector $[x(1), x(2), \cdots, x(N)]$ that contains

[1]For this lab we assume that we are making observations that are discrete in time, that is, t is integer.

all measurements/data from time 1 to time N. Then if an exact linear model can be found we would be able to find β_j as a solution to the system of equations:

$$\underbrace{\begin{bmatrix} x(B) & x(B-1) & \cdots & x(1) \\ x(B+1) & x(B) & \cdots & x(2) \\ x(B+2) & x(B+1) & \cdots & x(3) \\ \vdots & \ddots & \cdots & \vdots \\ x(N-1) & x(N-2) & \cdots & X(N-B) \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_B \end{bmatrix}}_{\boldsymbol{\beta}} = \underbrace{\begin{bmatrix} x(B+1) \\ x(B+2) \\ x(B+3) \\ \vdots \\ X(N) \end{bmatrix}}_{\mathbf{b}}, \tag{2}$$

where $\mathbf{A}$ is of size $(N-B) \times B$, and the vector $\mathbf{b}$ is of size $(N-B) \times 1$. As discussed in class, in most cases of interest the system

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{b} \tag{3}$$

does not have a solution because $N - B > B$, there are more equations than unknowns so that $\mathbf{A}$ is a very tall matrix and then it is unlikely that $\mathbf{b}$ will belong to the column space of $\mathbf{A}$. A standard approach to deal with this situation is instead find the least squares solution by solving instead:

$$\mathbf{A}^\top \mathbf{A}\boldsymbol{\beta} = \mathbf{A}^\top \mathbf{b}. \tag{4}$$

The system of (4) will always have a solution if $\mathbf{A}$ is full rank (i.e., $\text{rank}(\mathbf{A}) = B$). Notice that the matrix $\mathbf{A}^\top\mathbf{A}$ is square of size $B \times B$, and the vector $\mathbf{A}^\top\mathbf{b}$ has size $B \times 1$. The solution of (4) is called Least Squares solution, because it can be shown to be the minimizer of

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{B} (\mathbf{A}_i\boldsymbol{\beta} - b_i)^2 \tag{5}$$

where $\mathbf{A}_i$ is the i-th row of $\mathbf{A}$.

Problem 1. Suppose matrix $\mathbf{A}$ in (4) has full column rank. Find the explicit formula to compute the least squares solution: $\hat{\boldsymbol{\beta}}$. Which type of one-sided matrix inverse of $\mathbf{A}$ did you use to arrive at your answer?

Problem 2. Using the least squares solution: $\hat{\boldsymbol{\beta}}$, we obtain an estimate of $\mathbf{b}$: $\hat{\mathbf{b}} = \mathbf{A}\hat{\boldsymbol{\beta}}$. What is $\hat{\mathbf{b}}$ if $\mathbf{b} \in \mathbf{col}(\mathbf{A})$?

Problem 3. Show that $\hat{\mathbf{b}} - \mathbf{b} \in \mathbf{N}(\mathbf{A^T})$. hint: all least squares solutions must satisfy (4).

Problem 4. You are given temperature data[2] in the file data.mat, where the vector daily_temp stores daily average temperature (in degree Celsius) across the continental United States, from the years 2000 to 2016. The value daily_temp(1) corresponds to the average temperature of January 1st 2000, and the value daily_temp(end) the average temperature of December 31st 2016. The vector ndays contains the number of days of each year from 200-2016 (leap years have 366 days).

[2]https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.surface.html

1. Plot the vector daily_temp and describe what you observe. Include the plot in your report.

2. Implement a function that given a vector $[x(1), x(2), \cdots, x(N)]^\top$ and a parameter B, it returns the matrix $\mathbf{A}$ and the vector $\mathbf{b}$

3. You will estimate the linear model $\boldsymbol{\beta}$ using temperature from 2000-2015. Using the temperature data from 2000-2015, and $B = 7$, construct $\mathbf{A}$, $\mathbf{b}$, and solve (4) to find β.

4. Plot the true temperature $\mathbf{b}$, the predicted temperature $\mathbf{A}\boldsymbol{\beta}$, and the error vector $\mathbf{A}\boldsymbol{\beta} - \mathbf{b}$. During what part of the year is the prediction worst?

1.2 Temperature forecasting

In this section you will use the $\boldsymbol{\beta}$ estimated in the previous problem to predict the daily temperature during 2016. To do this, you will use the model computed in Problem 1 and apply it to predict temperatures in 2016 (note that these temperatures were not used to train the model). That is, for each interval of B consecutive days in 2016 you will predict the temperature of the next day, based on the model.

Problem 5. Using equation (1) do the following

1. Using $\boldsymbol{\beta}$ from Problem 1 (with $B = 7$) and equation (1) predict all temperatures for 2016. You can do this by constructing a matrix $\mathbf{C}$ and vector $\mathbf{d}$, such that $\mathbf{C}\boldsymbol{\beta}$ is the predicted temperature, and $\mathbf{d}$ is the true temperature of 2016.

2. Plot true temperature $\mathbf{d}$, predicted temperature $\mathbf{C}\boldsymbol{\beta}$ and error $\mathbf{C}\boldsymbol{\beta} - \mathbf{d}$.

Problem 6. Quantitative analysis: how does prediction accuracy change as a function of B?

- Find optimal $\boldsymbol{\beta}$ from Problem 1, for values of $B = \{1, 2, \cdots, 10\}$. Plot the prediction error for the norm[3] $|\mathbf{A}\boldsymbol{\beta} - \mathbf{b}|^2/(N - B)$ as a function of B. What is the best value of B?

- Apply the $\boldsymbol{\beta}$ from before, and compute prediction for 2016, and plot the error norm $|\mathbf{C}\boldsymbol{\beta} - \mathbf{d}|^2/366$ as a function of B. What is the best value of B?

- What happens to $\boldsymbol{\beta}$ as you increase B (besides the change of size)?, how does that affect the prediction?

[3] $|\mathbf{x}|^2 = \mathbf{x}^t\mathbf{x}$