



Machine Learning 2020

Homework 1

Due date : 2020/11/1 23:59:59 (Hard Deadline)

1 Bayesian Linear Regression

Given the training set \mathbf{x} and corresponding label set \mathbf{t} , we want to predict the label t of new test point x . In other words, we wish to evaluate the predictive distribution $p(t|x, \mathbf{x}, \mathbf{t})$.

A linear regression function can be expressed as below where the $\phi(x)$ is a basis function:

$$y(x, \mathbf{w}) = \mathbf{w}^T \phi(x)$$

In order to make prediction of t for new test data x from the learned \mathbf{w} , we will:

- Multiply the likelihood function of new data $p(t|x, \mathbf{w})$ and the posterior distribution of training set with label set.
- Take the integral over \mathbf{w} to find the predictive distribution:

$$p(t|x, \mathbf{x}, \mathbf{t}) = \int_{-\infty}^{\infty} p(t|x, \mathbf{w}) p(\mathbf{w}|\mathbf{x}, \mathbf{t}) d\mathbf{w}.$$

Now, please answer the following questions:

1. Why we need the basis function $\phi(x)$ for linear regression? And what is the benefit for applying basis function over linear regression? (5%)
2. Prove that the predictive distribution just mentioned is the same with the form

$$p(t|x, \mathbf{x}, \mathbf{t}) = \mathcal{N}(t|m(x), s^2(x))$$

where

$$m(x) = \beta \phi(x)^T \mathbf{S} \sum_{n=1}^N \phi(x_n) t_n$$
$$s^2(x) = \beta^{-1} + \phi(x)^T \mathbf{S} \phi(x).$$

Here, the matrix \mathbf{S}^{-1} is given by $\mathbf{S}^{-1} = \alpha \mathbf{I} + \beta \sum_{n=1}^N \phi(x_n) \phi(x_n)^T$ (15%)

(hint: $p(\mathbf{w}|\mathbf{x}, \mathbf{t}) \propto p(\mathbf{t}|\mathbf{x}, \mathbf{w}) p(\mathbf{w})$ and you may use the formulas shown in page 93.)

3. Could we use linear regression function for classification? Why or why not? Explain it! (10%)

2 Linear Regression

In this homework, you need to predict the chance of being admit in base on relevant student resume data. The following two approaches need to be realized respectively:

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



You are given a dataset ([dataset_X.csv](#), [dataset_T.csv](#)) to train your own linear regression model! Dataset provides total 500 students with 7 features. Can you use these features to predict the [chance of admit](#) for your own dream school?

One might consider the following steps to start the work:

1. Download and check for the dataset.
2. Create a new Colab or Jupyter notebook file.
3. Divide the dataset into training and validation. 0.2

Dataset Description

- [dataset_X.csv](#) contains 7 different resume feature served as the input.
GRE_score, TOFEL_score, University_rating, SOP, LOR, CGPA, Research
- [dataset_T.csv](#) contains the chance of admit regard as the target.
Chance_of_Admit

Specification

- For those problems with **Code Result** at the end, you must show your result in your .ipynb file or you will get **no** points.
- For those problem with **Explain** at the end, you must have a clear explanation or you will get low points.
- You are also encouraged to have some discussion on those problem which is not marked as **Explain**.

1. Feature select

In real-world applications, the dimension of data is usually more than one. In the training stage, please fit the data by applying a polynomial function of the form

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{i=1}^D w_i x_i + \sum_{i=1}^D \sum_{j=1}^D w_{ij} x_i x_j \quad (M = 2)$$

and minimizing the error function.

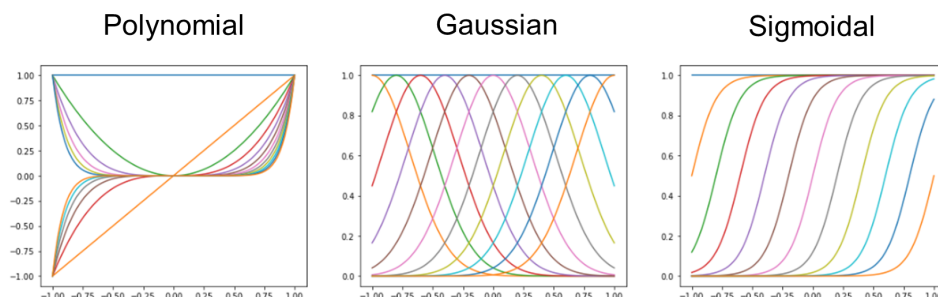
$$E(\mathbf{w}) = \frac{1}{2N} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2$$

- In the feature selection stage, please apply polynomials of order $M = 1$ and $M = 2$ over the dimension $D = 7$ input data. Please evaluate the corresponding RMS error on the training set and valid set. (15%) **Code Result**
- How will you analysis the weights of polynomial model $M = 1$ and select the most contributive feature? **Code Result, Explain** (10%)

2. Maximum likelihood approach

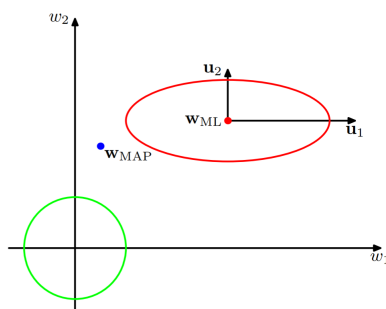
- Which basis function will you use to further improve your regression model, Polynomial, Gaussian, Sigmoidal, or hybrid? **Explain** (5%)
- Introduce the basis function you just decided in (a) to linear regression model and analyze the result you get. (Hint: You might want to discuss about the phenomenon when model becomes too complex.) **Code Result, Explain** (10%)

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



- Apply N-fold cross-validation in your training stage to select at least one hyper-parameter (order, parameter number, ...) for model and do some discussion (underfitting, overfitting). **Code Result, Explain** (10%)

3. Maximum a posterior approach



- (a) What is the key difference between maximum likelihood approach and maximum a posterior approach? **Explain** (5%)
- (b) Use Maximum a posterior approach method to retest the model in **2** you designed. You could choose Gaussian distribution as a prior. **Code Result** (10%)
- (c) Compare the result between maximum likelihood approach and maximum a posterior approach. Is it consistent with your conclusion in (a)? **Explain** (5%)

3 Rules

- Please name the assignment as **hw1_StudentID.zip** (e.g. hw1_0123456.zip).
- In your submission, it needs to contain three files.
 - **.ipynb** file which contains all the results and codes for this homework.
 - **.py** file which is downloaded from the .ipynb file
 - **.pdf** file which is the report that contains your description for this homework.
- Implementation will be graded by
 - Completeness
 - Algorithm Correctness
 - Model description
 - Discussion
- Only **Python** implementation is acceptable.
- **Numpy** library is recommended for the implementation.
- **Don't use high level toolbox/module functions**(e.g. sklearn, polyfit).
- **DO NOT PLAGIARISM.** (We will check program similarity score.)