Statistical Modelling & Machine Learning HW1

(Due: 10/10/2021, Sunday)

1. Suppose that Y_i, $i = 1, \ldots, n$, are independent and $\mu_i = E(Y_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, where $\boldsymbol{x}_i$ is the ith input vector and $\boldsymbol{\beta}$ is an unknown parameter vector. Suppose that Y_i's have double exponential distribution as follows:

$$p_{Y_i}(y_i; \mu_i, \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|y_i - \mu_i|}{\sigma} \right).$$

For the fixed σ, Show that the maximum likelihood estimate of $\boldsymbol{\beta}$ can be obtained by minimizing $\sum_{i=1}^{n} |y_i - \hat{y}_i|$, where $\hat{y}_i$ satisfies the linear model.

2. Consider the linear model as follows:

$$Y_t = \boldsymbol{X}_t^\top \boldsymbol{\beta} + \epsilon_t, \;\; t = 1, \ldots, T,$$

where $\epsilon_t = e_t + \theta e_{t-1}$, where $|\theta| < 1$. Here we assume $\boldsymbol{e}_t \sim^{iid} N(0, \sigma^2)$ for all t. Specify the covariance matrix of $\boldsymbol{Y} = (Y_1, \ldots, Y_T)^\top$.

3. Consider the spatial autoregressive model. Suppose that we collected 4 observations from the 2 dimensional coordinate (z_1, z_2) and we use the exponential distance weight with $\alpha = 1$ and the Euclidean distance for d_{ij} to construct the spatial weight matrix. Find the spatial weight matrix. Also, discuss the spatial correlation structure of the data.

	z_1	z_2
Obs 1	5	3
Obs 2	1	4
Obs 3	2	2
Obs 4	4	1

4. Suppose that $Y_i \sim Poisson(\mu_i)$, $i = 1, \ldots, n$ and Y_i's are independent.

(1) Show that the Poisson distribution belongs to the exponential family.

(2) Find the canonical link function.

5. Consider '`Q5.csv`' data file. To predict the variable `Y` based on `(X1, X2, X3)`, use a **data modelling technique**.

(1) Investigate whether there is an irrelevant input variable for the prediction of `Y`. If it exists, find it and justify why it is the irrelevant variable.

(2) Construct the best parametric regression model and estimate the model parameters.

(3) Show the residual plot for the best model obtained from part (2).

(4) Based on part (2), describe the functional relationships between `Y` and individual input variables in the model.